HAL
open science

Gestion du résultat et seuils comptables : Impact des choix méthodologiques et proposition d'un instrument de mesure des irrégularités.

Olivier Vidal

To cite this version:

Olivier Vidal. Gestion du résultat et seuils comptables : Impact des choix méthodologiques et proposition d'un instrument de mesure des irrégularités.. Gestion et management. HEC PARIS, 2008. Français. NNT : . tel-00483265

HAL Id: tel-00483265

https://pastel.hal.science/tel-00483265

Submitted on 13 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ecole Doctorale n° 471
« SCIENCES DE LA DECISION ET DE L'ORGANISATION »

ECOLE DES HAUTES ETUDES COMMERCIALES DE PARIS

Ecole Doctorale « Sciences de la Décision et de l'Organisation » - ED 471
Equipe de Recherche GREGHEC - UMR 2959

« GESTION DU RESULTAT ET SEUILS COMPTABLES : IMPACT DES CHOIX METHODOLOGIQUES ET PROPOSITION D'UN INSTRUMENT DE MESURE DES IRREGULARITES »

THESE
présentée et soutenue publiquement le 3 décembre 2008
en vue de l'obtention du
DOCTORAT EN SCIENCES DE GESTION
par

Olivier VIDAL

JURY

Président du Jury : **Monsieur Patrick BOISSELIER**
Professeur des Universités
Institut d'Administration des Entreprises de Nice

Directeur de recherche : **Monsieur Hervé STOLOWY**
Professeur
Ecole des Hautes Etudes Commerciales

Rapporteurs : **Monsieur François DEGEORGE**
Professeur
Université de Lugano - Suisse

Monsieur Charles PIOT
Professeur des Universités
Université Pierre Mendès-France - Grenoble

Suffragants : **Madame Christine POCHET**
Professeur des Universités
Institut d'Administration des Entreprises de Paris

Monsieur Thomas JEANJEAN
Professeur-Associé
Ecole des Hautes Etudes Commerciales

École des Hautes Études Commerciales

Le Groupe HEC Paris n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Remerciements

Même si une thèse est le résultat d'un cheminement personnel, elle ne peut aboutir sans puiser dans les apports multiples d'un grand nombre de personnes. Mes remerciements s'adressent donc à tous ceux qui, directement ou indirectement, m'ont aidé, orienté et soutenu.

En premier lieu, je tiens à remercier le Professeur Hervé Stolowy pour la confiance qu'il a témoignée en acceptant de diriger ma recherche. C'est grâce à lui que j'ai eu le courage de me lancer dans cette aventure. Son soutien indéfectible, ses conseils et sa rigueur m'ont accompagné tout au long de ces quatre années de travail, et m'ont beaucoup apporté.

Je remercie les Professeurs François Degeorge et Charles Piot d'avoir accepté d'être mes rapporteurs, ainsi que les Professeurs Patrick Boisselier, Thomas Jeanjean et Christine Pochet de me faire l'honneur de participer à mon jury.

Je remercie le Doctorat HEC et plus précisément tous les membres du département ICCO pour la qualité de l'encadrement et des ressources dont j'ai pu bénéficier au cours de ma thèse. Je remercie également l'ENS Cachan et mes collègues du département Économie-Gestion pour leur soutien, et plus spécifiquement Nicolas Drouhin, Philippe Fontaine, Dominique Guégan et Éric Savaterro qui, en aménageant ma charge d'enseignement, ont permis la réalisation de mes travaux. Je remercie enfin l'ESSEC pour m'avoir accueilli sur son campus de Singapour durant près de six mois lors de la rédaction de ce document.

Je remercie tous ceux qui ont contribué à faire avancer ma réflexion, à l'occasion de séminaires ou de présentations. Isabelle Chambost et les participants des séminaires doctoraux du CNAM m'ont permis d'échanger idées et expériences. Henry Yue, en m'invitant à la Guanghua School of Management (Université de Pékin), a contribué grandement à l'enrichissement de ma recherche. Marc Lassagne et les doctorants du GRID (ENSAM) ont élargi ma réflexion sur la nature des seuils comptables. Yuan Ding et Pascal Dumontier ont, à l'occasion des présentations de mes travaux en séminaires ou lors de conférences, significativement participé à orienter mes travaux.

Je remercie toutes les personnes qui m'ont apporté une aide précieuse, parfois ponctuelle, souvent durable, toujours généreuse. Parmi elles, Yves Mard et Gwénaëlle Nogatchewsky m'ont conseillé et éclairé à de nombreuses reprises. Ganaël Bascoul et Jean-Christophe Tavanti m'ont particulièrement aidé à construire le modèle paramétrique de mesure des irrégularités.

Je remercie sincèrement Sophie Audousset et tous mes compagnons de route, Marion Brivot, Marc Debourmont, Sophie Marmouset et Emmanuelle Plot, pour leur présence et toutes les discussions que nous avons pu avoir. Je conserve par ailleurs un excellent souvenir de l'ambiance vraiment chaleureuse de la promotion 2003 du doctorat HEC ; les liens de solidarité qui se sont tissés à cette occasion ont été sans conteste une grande force pour tous dans notre travail.

Enfin, je réserve mes remerciements les plus chaleureux à Carole sans qui le projet de me lancer dans cette recherche n'aurait pas pu voir le jour.

Plan simplifié de la thèse

Remerciements 3
Plan simplifié de la thèse 4

INTRODUCTION GÉNÉRALE 6

PARTIE 1 : DÉFINITION DU CHAMP DE RECHERCHE 9

Chapitre I - Réflexion sur la valeur des informations comptables 10
A. L'origine de la question de recherche 10
B. Le cadre général dans lequel s'insère la recherche 12

Chapitre II - Littérature sur les seuils 28
A. Les travaux observant les seuils 28
B. Les travaux sur les causes des seuils 32
C. Les travaux étudiant les moyens et les conséquences 49
D. Les travaux mesurant et comparant les seuils 55
E. Synthèse de la revue de littérature et définition des axes de recherche 60

Chapitre III - Cadre théorique explicatif de la gestion des seuils 66
A. Un éclairage terminologique conduisant à distinguer deux « effets de seuils » 67
B. Proposition d'un cadre d'analyse 75
C. Conclusion de la partie 1 : problématique de la thèse 82

PARTIE 2 : IMPACT DES CHOIX MÉTHODOLOGIQUES 84

Chapitre IV – État des lieux et présentation de l'étude méthodologique 85
A. Impact de la méthode de mesure de l'irrégularité 86
B. Impact de la largeur des intervalles d'observation 115
C. Impact de la variable de mise à l'échelle 117

Chapitre V – Démarche générale de l'étude empirique et analyse de la base de données 124
A. Analyse préliminaire de la base initiale 124
B. Correction de la base 129
C. Analyse descriptive de la base corrigée 131

Chapitre VI – Mise en œuvre de l'étude méthodologique 135

A. Impact de la méthode de mesure de l'irrégularité.....135

B. Impact de la largeur des intervalles, et nouvel indicateur d'irrégularité.....177

C. Impact du choix de la variable de mise à l'échelle.....198

D. Conclusion de la partie 2 : Limites des mesures non paramétriques.....234

PARTIE 3 : PROPOSITION D'UN INSTRUMENT DE MESURE ET APPLICATION AUX ENTREPRISES FRANÇAISES COTÉES 237

Chapitre VII - Proposition d'un nouvel instrument de mesure des irregularités 238

A. Pourquoi faut-il adopter une méthode paramétrique ?.....238

B. Existe-t-il une loi de distribution des résultats ?.....241

C. À la recherche d'une loi de distribution 250

Chapitre VIII - Mise en œuvre : ajustement d'une loi de distribution des résultats des entreprises françaises..... 262

A. Valeurs extrêmes.....262

B. Réduction.....266

C. Centrage.....267

D. Ajustement.....270

Chapitre IX - Application dans le cadre d'une étude longitudinale 303

A. Étude longitudinale de l'évolution des seuils en France 303

B. Deux nouveaux indicateurs d'irrégularité.....316

C. Conclusion de la partie 3 : Les seuils comptables en France 347

CONCLUSION GÉNÉRALE.....351

ANNEXES.....357

BIBLIOGRAPHIE ET INDEX DES FIGURES ET TABLEAUX.....457

Bibliographie 457

Index des figures.....464

Index des tableaux 469

TABLE DES MATIÈRES DÉTAILLÉE 475

INTRODUCTION GÉNÉRALE

Les recherches sur les seuils comptables s'intègrent dans le cadre général de la qualité des informations financières, cadre qui résonne avec l'actualité depuis les scandales financiers de la fin du vingtième et du début du vingt-et-unième siècle d'une part, et l'accélération du processus d'harmonisation internationale des normes comptables d'autre part.

S'attendant à ce que la distribution statistique des résultats des entreprises présente une allure gaussienne, plusieurs chercheurs ont interprété les irrégularités de distribution des résultats publiés par les entreprises comme la manifestation de manipulations comptables pour atteindre des seuils. Les travaux sur les seuils se sont développés rapidement à partir de 1997, complétant ainsi les travaux sur la gestion du résultat utilisant les modèles plus traditionnels d'accruals anormaux développés dans les années 1980.

Les travaux sur les seuils ont identifié différents types de seuils, proposé des facteurs explicatifs à ces seuils, et suggéré certaines de leurs conséquences sur la performance future des entreprises. Si leurs apports sont importants pour la recherche en comptabilité, des limites méthodologiques ont été identifiées, rendant les mesures effectuées jusqu'alors difficiles à interpréter.

En étudiant leur évolution en France, l'objectif de la thèse est de mieux comprendre le phénomène des seuils comptables. Au-delà du seul contexte français, la thèse propose un outil méthodologique permettant d'améliorer les modèles d'identification, de mesure et de comparaison des irrégularités de distribution des résultats comptables. Ces modèles étaient jusqu'alors essentiellement fondés sur une approche non paramétrique des irrégularités de distribution. La thèse met en œuvre un modèle paramétrique, ce qui suppose donc de connaître la loi de distribution théorique des résultats publiés. Le modèle a l'originalité de s'appuyer sur une démarche empirique pour ajuster au mieux cette loi économique.

Le modèle est ensuite mis en application dans le cadre d'une étude longitudinale sur un échantillon d'entreprises françaises cotées. Il apparaît dès lors que près de 5 à 10 % de l'ensemble des entreprises cotées françaises semblent manipuler leurs comptes pour publier un résultat faiblement positif. Ce nombre tend à diminuer, essentiellement depuis 2001. Par ailleurs, la démarche paramétrique permettant d'estimer les montants potentiellement manipulés par ces entreprises, il apparaît que l'hypothèse de

manipulation marginale pour éviter de publier une légère perte n'est pas vérifiée. Autrement dit, au-delà du seul contexte français, l'étude tend à montrer que les entreprises qui évitent de publier une perte pour publier un petit bénéfice ne manipulent pas nécessairement des montants faibles. En élargissant la portée de cette conclusion, la thèse tend à renforcer l'importance de la gestion pour atteindre des seuils, et souligne l'intérêt de poursuivre les recherches sur ce thème.

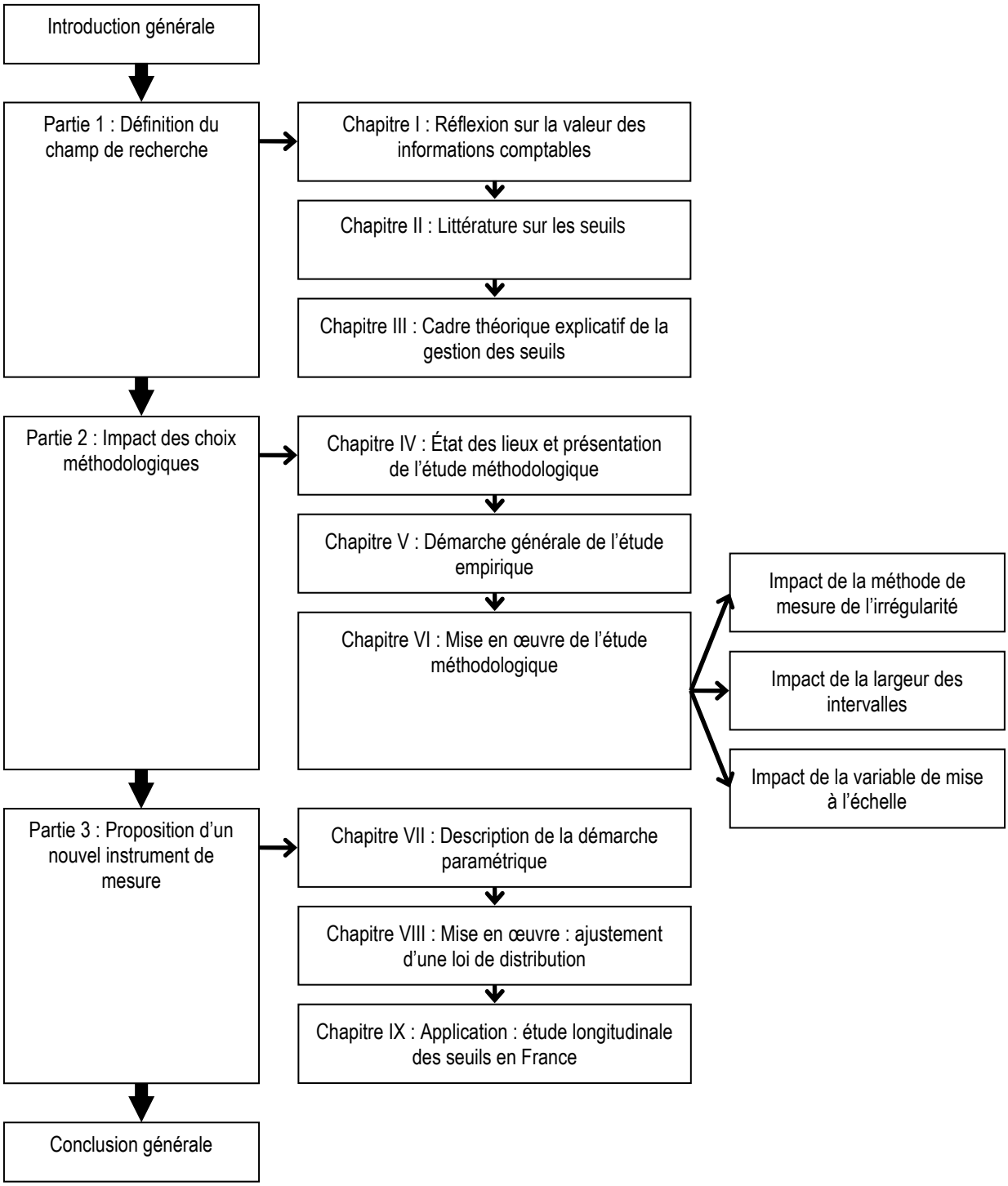
La thèse est structurée en 3 parties. La première partie définit le champ de la recherche et le situe dans le panorama général de la recherche en comptabilité. Une attention particulière est apportée aux explications proposées pour justifier l'existence des seuils comptables. La revue de littérature débouche sur un double constat : (1) Il manque un cadre conceptuel permettant d'analyser les seuils comptables, et (2) des limites méthodologiques apparaissent dans les recherches antérieures. Ce double constat permet alors de préciser le sujet de la thèse : au-delà de l'étude des seuils en France, et de leur évolution, l'ambition de la thèse est d'apporter une réponse aux limites méthodologiques identifiées. Le cadre d'analyse proposé en fin de partie 1, unifiant les explications comportementales avec le cadre de la recherche positive en comptabilité, répond ainsi au premier constat. Le second constat (limites méthodologiques) fait l'objet des parties 2 et 3 de la thèse.

La seconde partie précise les limites méthodologiques identifiées dans la littérature. Leurs conséquences sont étudiées à travers une application empirique auprès de l'échantillon d'entreprises cotées françaises. Ces limites méthodologiques ont peu d'impact lorsqu'il s'agit de mettre en évidence les irrégularités qui sont manifestes. Mais elles ne peuvent être ignorées lorsqu'il s'agit de mesurer et comparer les irrégularités, dans le temps ou dans l'espace. Plus précisément, la partie 2 de la thèse examine ainsi successivement l'impact du choix de la méthode retenue pour mesurer les irrégularités, l'impact du choix de la largeur de l'intervalle d'étude, et l'impact du choix du dénominateur. Cette partie de l'étude montre que si les études sont si sensibles aux choix méthodologiques, cela est essentiellement dû à la nature non paramétrique des démarches mises en place jusqu'alors pour étudier les seuils comptables.

La partie 3 tente de répondre à ces limites en présentant un outil original de mesure fondé sur une démarche paramétrique. Une telle démarche suppose de connaître la loi de distribution théorique des résultats publiés (fonction mathématique). Cette loi étant jusqu'alors inconnue, une approche empirique est proposée pour l'ajuster. La démarche est alors mise en application sur la base de populations d'entreprises françaises, allemandes et britanniques, cotées de 1992 à 2004. Elle permet de sortir du dilemme du

choix de la mesure. Elle permet surtout de mesurer les irrégularités sur l'ensemble de la distribution, notamment en élargissant les intervalles d'étude, ce qui autorise une estimation du nombre total d'entreprises manipulatrices, et des montants manipulés. La démarche est alors mise en œuvre pour mesurer l'évolution des irrégularités de distribution des entreprises françaises de 1992 à 2004.

L'articulation de ces trois parties est synthétisée dans la figure ci dessous :



PARTIE 1 : DÉFINITION DU CHAMP DE RECHERCHE

La première partie de la thèse est essentiellement théorique. Elle s'oppose en cela aux deux parties empiriques suivantes.

Les deux premiers chapitres sont construits comme un entonnoir. Partant d'un constat initial, le chapitre I situe d'abord le questionnement originel dans le cadre général de la recherche en comptabilité, et plus particulièrement dans celui de la littérature sur la gestion du résultat. L'étude des seuils comptables est apparue en effet comme une alternative aux mesures d'accruals pour étudier la gestion du résultat. Elle est à l'origine d'un nombre important et croissant de publications ces dernières années.

Le chapitre II brosse un panorama de la littérature sur les seuils comptables, en distinguant les études qui identifient les seuils, celles qui les mesurent et les comparent, et celles qui étudient leurs conséquences. Une attention particulière est apportée aux explications proposées pour justifier l'existence des seuils comptables. La revue de littérature débouche sur un double constat : (1) Il manque un cadre conceptuel permettant d'analyser les seuils comptables, et (2) des limites méthodologiques apparaissent dans les recherches antérieures.

Ce double constat permet alors de préciser le sujet de la thèse : au-delà de l'étude des seuils en France, et de leur évolution, l'ambition de la thèse est d'apporter une réponse aux limites méthodologiques identifiées. Le chapitre III répond au premier constat. Une réflexion sémantique, précisant notamment la notion d'effet de seuil et proposant une typologie des seuils, débouche sur la proposition d'un cadre d'analyse unifiant les explications comportementales avec le cadre de la recherche positive en comptabilité.

Le second constat (limites méthodologiques) fait l'objet des parties 2 et 3 de la thèse.

CHAPITRE I - RÉFLEXION SUR LA VALEUR DES INFORMATIONS COMPTABLES

Le chapitre I s'ouvre sur un constat initial (section A) : l'observation graphique d'irrégularités dans la distribution des résultats des entreprises françaises. Ce constat suscite des questions quant à leur interprétation. Ces questions sont à l'origine le l'étude. Mais le sujet de thèse, et notamment son orientation méthodologique ne seront présentés qu'après la revue de littérature (chapitre II). La section B situe ces questions dans le cadre plus général de la recherche en comptabilité.

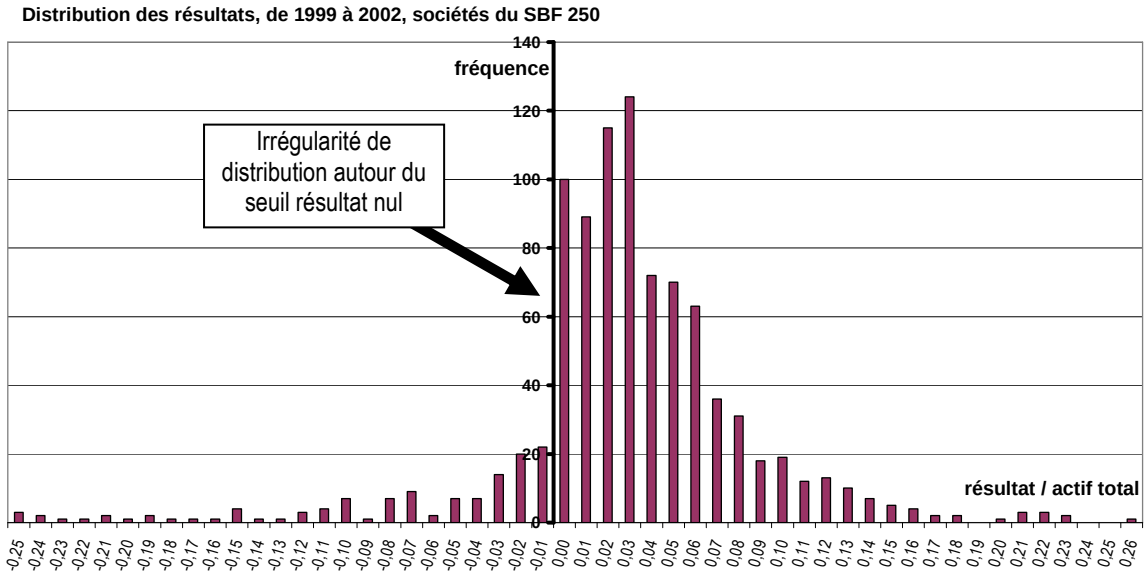
A. L'origine de la question de recherche

1. Un constat initial…

On pourrait s'attendre à ce que la distribution statistique des résultats des entreprises présente une allure « plus ou moins » gaussienne. Or plusieurs auteurs ont constaté des irrégularités de distribution des résultats publiés. Au cours du mois de juillet 2004, une étude exploratoire a précédé le travail de thèse afin de cerner la faisabilité d'une recherche sur le thème des seuils comptables. L'étude portait sur des entreprises françaises cotées incluses dans l'indice SBF 250. Les données comptables des entreprises de 1999 à 2002 sont extraites de la base Global (Standard & Poor's).

Comme l'illustre la Figure 1, cette étude permet de vérifier les résultats des études antérieures (Burgstahler et Dichev, 1997 ; Degeorge, Patel et Zeckhauser, 1999) : des irrégularités de distribution autour des seuils « résultat nul » et « variation nulle de résultat » sont constatées, et plusieurs pistes de recherches sont alors suggérées.

Figure 1 : Étude exploratoire. Constatation des irrégularités de distribution des résultats de 1999 à 2002 par les entreprises du SBF 250



Le plan de la thèse ne suit pas, on s'en doute, le parcours chronologique du travail effectué par le chercheur. Mais cette accroche « historique » a pour mission de mettre en relief l'étonnement initial, cette étincelle qui éveille la curiosité, puis alimente l'intérêt du chercheur tout au long de son travail.

2. …qui suscite des questions

Comment se fait-il que les distributions de résultats publiés par les entreprises présentent des « pics » et des « creux » ? Peut-on justifier ces irrégularités mathématiquement ? Comment les mesurer, les comparer, les interpréter ?

Contrairement à la tendance mondiale, le thème de la gestion du résultat pour atteindre des seuils a donné lieu à très peu de recherches en France. Comment la gestion du résultat pour atteindre les seuils se manifeste-t-elle en France ? Quelle est son évolution ? Comment la comparer aux pratiques déjà décelées dans d'autres pays ? Telles sont les questions spontanées qui sont à l'origine de la rédaction de la présente thèse.

Pour répondre à ces questions, il est indispensable d'identifier et surtout de mesurer les irrégularités qui se manifestent en France. La première formulation générale du sujet conduit donc à proposer la démarche de recherche suivante :

Une revue de littérature doit (1) préciser la question de recherche, et (2) définir une méthode de mesure des irrégularités adaptée à l'objectif de comparaison fixé. Les mesures pourront alors être effectuées sur un échantillon d'entreprises françaises.

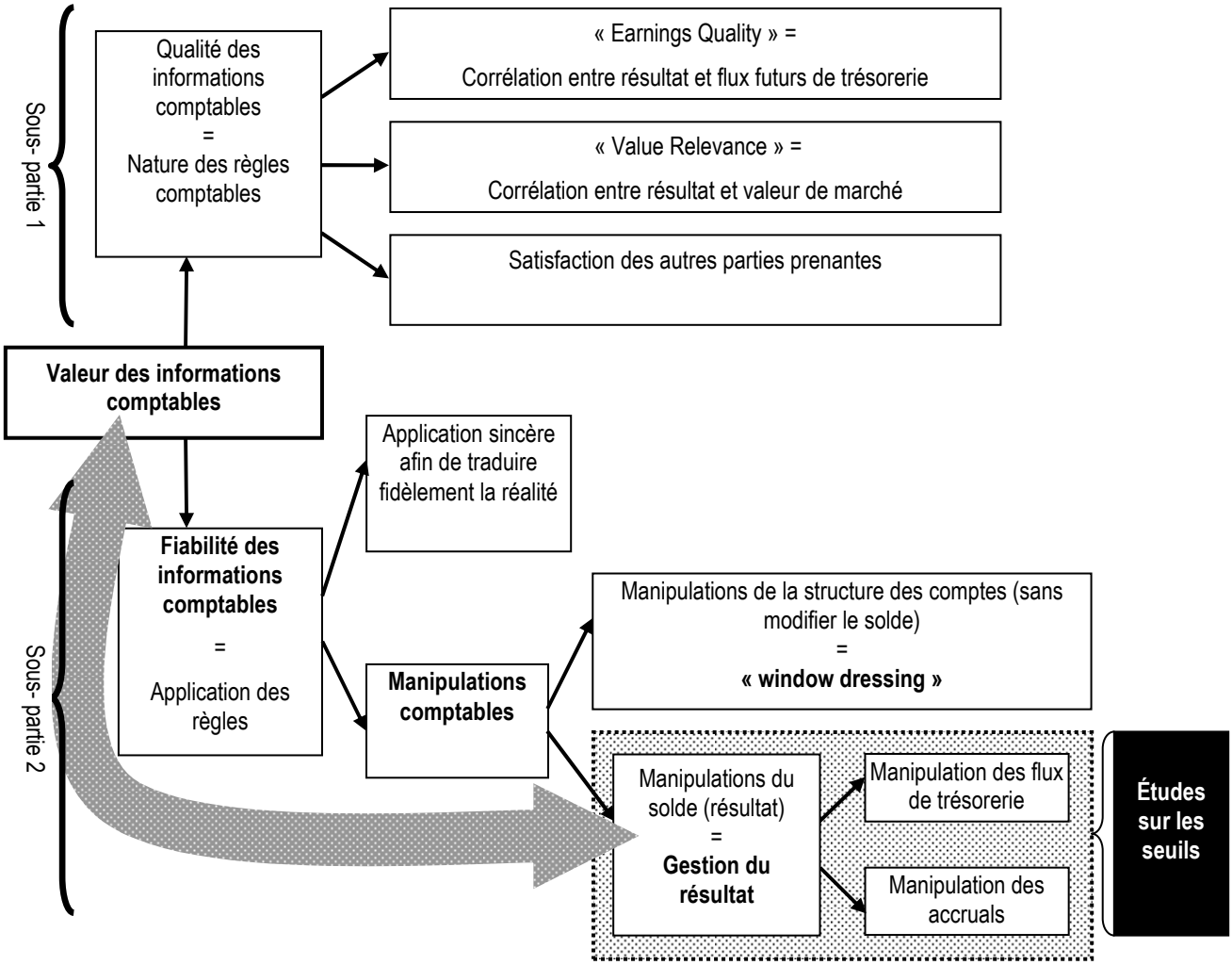
Mais la revue de littérature révèle des surprises qui conduiront à redéfinir la question de recherche dans la section E du chapitre II. La thèse prendra alors une orientation beaucoup plus méthodologique.

B. Le cadre général dans lequel s'insère la recherche

Les premiers articles à s'être intéressés aux seuils comptables, en identifiant des discontinuités de distribution des résultats publiés, ont soulevé tout naturellement des questions sur la valeur informative des données comptables publiées, sur leur qualité et leur fiabilité. Puisque les états financiers fournissent une information qui permet la prise de décisions, la comptabilité est-elle à même de répondre aux attentes de ses utilisateurs ?

La question de la valeur des informations comptables (et notamment du résultat) a naturellement nourri un très important courant de recherches en comptabilité. Il est possible de distinguer deux approches complémentaires pour aborder cette question. (1) Les travaux de recherche sur la « qualité » s'intéressent à la pertinence des règles comptables d'une part, et (2) les travaux sur la « fiabilité » identifient d'éventuelles manipulations comptables. C'est dans ce deuxième courant que s'intègre le champ d'étude sur les seuils comme l'illustre la Figure 2.

Figure 2 : La question de recherche dans le cadre plus général de la recherche en comptabilité sur la valeur des informations comptables



La sous-partie 1 (ci-dessous) rappelle les problématiques liées à la qualité des informations comptables. La sous-partie 2 développe plus largement les problématiques de recherche sur la fiabilité dans lesquelles s'intègrent les questions sur les seuils. La littérature sur les seuils est développée dans le chapitre II.

1. La qualité des informations comptables

a. Résultat et qualité

L'essence même de la comptabilité d'engagement (« accrual accounting ») est d'enregistrer les flux, non pas lorsqu'ils ont réellement lieu, mais lorsque la règle comptable le décide. C'est la différence entre la « comptabilité » en tant que discipline, et

la « comptabilisation » en tant que technique d'enregistrement et de calcul. Sans ces décalages d'enregistrement, que les anglo-saxons nomment « accruals », le résultat comptable ne serait rien d'autre qu'un solde de trésorerie (comptabilité de caisse).

Le terme anglais « accruals » seul n'a pas actuellement de traduction en langue française. Il n'est pas de tradition de le traduire par le terme « engagements » ou « décalages »[1]. Seule l'expression complète « accrual accounting » est traduite par l'expression française « comptabilité d'engagement ». Tout se passe comme si en français, le concept même d'accruals n'existait pas. Le terme accruals est issu du verbe « to accrue » qui signifie « accumuler ». Les accruals correspondent donc conceptuellement à une « accumulation », une sorte de « matelas » de résultat. Ce terme est généralement traduit par l'expression[2] : « produits et charges calculés » (Dumontier et Raffournier, 1999). Dans le cadre de la thèse, aucune traduction n'est proposée, et le terme anglais « accruals » est utilisé.

En définitive, la règle comptable poursuit une finalité propre : donner au résultat une information spécifique que la seule « comptabilisation » des flux réels n'a pas. C'est sa « qualité », sa raison d'être. Le résultat donne (il tend à donner) une « image fidèle », c'est-à-dire une information synthétique sur la santé et la pérennité de l'entreprise qu'aucun autre indicateur ne peut fournir.

Le terme de « qualité » comptable étant à géométrie variable, il est nécessaire à ce stade de l'étude, d'en préciser sa définition, ou plus précisément, ses définitions. Le terme « qualité » peut être employé dans trois sens différents qui correspondent à trois niveaux de généralisation différents. Jusqu'alors, dans le cadre de l'introduction, le terme « qualité » est utilisé dans son sens courant, c'est-à-dire le plus large, évoquant la capacité qu'a l'information comptable à répondre aux besoins de ses utilisateurs. Les normes ISO définissent la qualité comme un ensemble de caractéristiques qui confèrent l'aptitude à satisfaire à des besoins[3].

Dans la suite de ce chapitre, un sens plus réduit du terme « qualité » est privilégié, permettant de l'opposer à celui de « fiabilité ». L'objectif est de distinguer la qualité propre

[1] Par opposition aux flux « réels » de trésorerie, les accruals sont des flux comptables n'ayant pas nécessairement de consistance monétaire. Ce sont des charges et produits calculés selon « l'arbitraire » des règles comptables, c'est-à-dire les variations de besoin en fond de roulement et les autres charges et produits calculés.

[2] Au Québec, l'expression « variables comptables de régularisation » est également utilisée.

[3] La définition ISO 9000 a évolué entre 1982 (« aptitude d'un produit ou d'un service à satisfaire au moindre coût et dans les moindres délais les besoins des utilisateurs ») et 2000 (« aptitude d'un ensemble de caractéristiques intrinsèques à satisfaire des exigences »). La définition retenue est une version simplifiée de l'expression de 1994 : « ensemble de caractéristiques d'une entité qui lui confère l'aptitude à satisfaire des besoins exprimés et implicites »).

(intrinsèque) aux règles et usages comptables, des difficultés liées à leur bonne application (notion de fiabilité de l'information). Ainsi, des règles peuvent être de qualité, mais leur mauvaise application et leur contournement plus ou moins volontaires par les acteurs économiques peuvent nuire à la fiabilité des informations comptables publiées.

Enfin, un troisième sens plus réducteur fait écho au champ de littérature anglo-saxon dénommé « earnings quality » qui limite l'étude de la qualité du résultat à sa capacité à prédire les flux de trésorerie futurs de l'entreprise. Ce troisième sens, très réducteur, est rapidement abordé dans le paragraphe b.1. de la page suivante, mais il n'est ensuite pas retenu dans le cadre de la thèse.

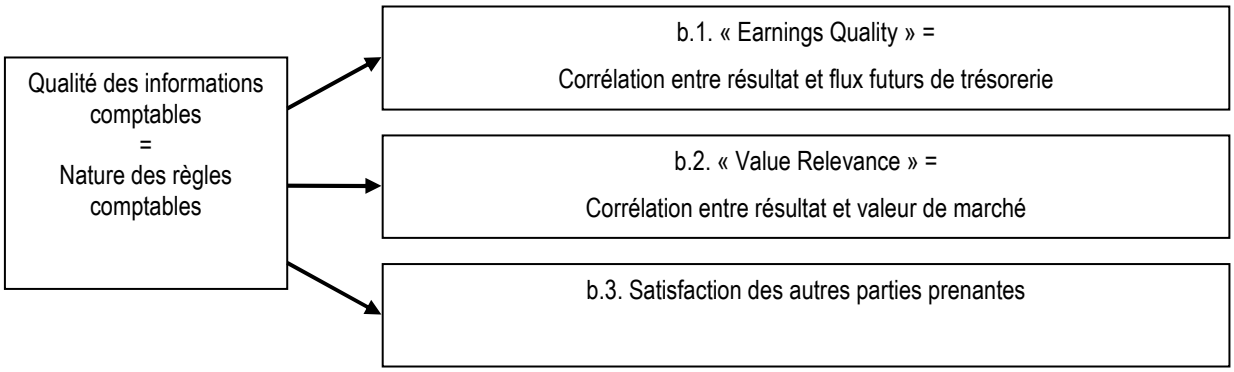
En conclusion, puisque la notion de qualité se définit en rapport avec un référentiel (les besoins et les attentes des utilisateurs du bien ou du service), il en découle que pour mesurer la qualité du résultat, il est indispensable de définir un indicateur (le référentiel) mesurant les besoins des utilisateurs de la comptabilité. C'est un point délicat pour la recherche en comptabilité, développé dans la section suivante.

b. Mesure de la qualité du résultat comptable

Si le besoin de l'utilisateur de la comptabilité est d'obtenir une image fidèle de la santé et la pérennité de l'entreprise, la difficulté à laquelle ont été confrontés les chercheurs a été d'en trouver un indicateur quantifiable. Comment définir un référentiel permettant de comparer le résultat mesuré par la comptabilité à cette image fidèle recherchée ? La tâche est difficile, voire impossible, dans la mesure où l'image fidèle est un concept théorique, qui ne se mesure que par le résultat ! C'est la même chose que le QI qui ne mesure pas l'intelligence, mais la capacité à répondre aux questions censées mesurer ce même QI.

Pour sortir de ce paradoxe, les travaux qui tentent de mesurer la qualité du résultat ont proposé deux modèles (schématisés dans la Figure 3) pour formaliser les besoins des actionnaires : l'approche par les cash flows (les flux de trésorerie) et l'approche par les cours boursiers (le marché). Une troisième approche fondée sur la satisfaction des utilisateurs peut être évoquée même si elle ne fait pas l'objet d'applications empiriques.

Figure 3 : La qualité des informations comptables



b.1. L'approche par les flux de trésorerie

L'idée est que le résultat s'écarte des enregistrements de trésorerie pour mieux prévoir les flux futurs. On étale les charges afin de les relier aux encaissements futurs qu'elles vont générer. On anticipe ainsi les entrées et les sorties futures.

Dans cette approche, la mesure de la qualité du résultat se réalise en comparant le résultat comptable d'une année et les flux de trésorerie des années suivantes. La littérature correspondant à cette approche dans la littérature anglo-saxonne porte le nom de « earnings quality » (Jones, 1991 ; Barth, Nelson et Cram, 2001 ; Dechow et Dichev, 2002 ; Cohen, 2003). Elle correspond au troisième sens, le plus strict, de l'expression « qualité comptable » évoquée plus haut (sous partie a.).

Cette approche souffre cependant de deux limites. Tout d'abord, les flux peuvent être eux-mêmes manipulés (modification des délais de paiement, retard dans le lancement d'une campagne de communication ou d'un investissement…). Les considérer comme étalon peut donc être source d'erreur. Mais surtout, et par définition, les flux futurs sont incertains. Le résultat comptable qui cherche à s'en approcher ne pourra jamais être parfaitement prédictif (et un résultat parfaitement prédictif serait fortement soupçonné d'être manipulé). Il y a là une rationalisation a posteriori qui limite conceptuellement la pertinence de cette approche.

b.2. L'approche par le marché

Cette approche suppose que le marché (les actionnaires) est efficient, et que le prix des actions intègre la totalité de l'information nécessaire aux investisseurs à un instant « t ». Dès lors, si le résultat est de qualité pour les actionnaires, c'est qu'il fournit une

information corrélée au marché puisque répondant au mieux aux attentes des actionnaires.

La mesure de la qualité du résultat se résume alors à chercher si la comptabilité enregistre à temps les (bonnes et/ou mauvaises) informations[4]. Dans la littérature anglo-saxonne, de telles études mesurant la corrélation entre un cours boursier et un résultat comptable se placent sous l'intitulé général des études sur la « value relevance » du résultat comptable (Ball et Brown, 1968 ; Beaver, 1968 ; Dumontier et Labelle, 1998). L'expression « value relevance » peut être traduite par le terme « pertinence » du résultat (Dumontier et Raffournier, 1999). Mais il faut souligner que la pertinence est limitée à une optique financière, supposant l'efficience des marchés : la pertinence n'est vue que du seul point de vue des propriétaires (actionnaires).

Cette approche souffre également d'une limite conceptuelle. Si un résultat comptable de qualité est celui qui est proche du cours, alors quel est l'intérêt de mesurer un résultat comptable distinct du cours boursier ? Le cours ne serait-il pas le meilleur indicateur de la santé de l'entreprise ?

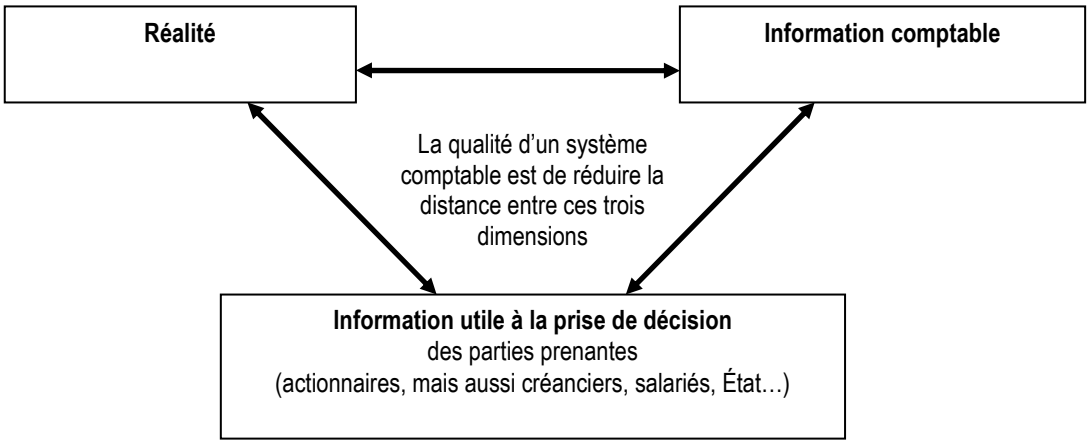
b.3. Autres approches

Les deux approches précédentes (mais surtout l'approche par le marché) mettent au cœur de leurs préoccupations la vision des propriétaires (actionnaires). D'autres approches sont envisageables, mettant en avant les besoins des autres parties prenantes. On peut ainsi se poser les questions suivantes : Le résultat est-il pertinent pour les créanciers ? Est-il pertinent pour l'État ? Pour les salariés ?... Ces pistes ne font pas, à notre connaissance, l'objet de recherches empiriques. Ce point, malgré son intérêt, n'a pas vocation à être développé dans le cadre de la thèse.

En définitive, puisque l'information comptable est une traduction de la réalité de l'entreprise, un système comptable de qualité est un système qui fournit à ses utilisateurs une information répondant à ses besoins, tout en étant fidèle à la réalité représentée. C'est ce qu'illustre la Figure 4.

[4] Notion de « timeliness » en anglais.

Figure 4 : Le filtre de l'information comptable



Toutes les problématiques de qualité développées ci-dessus reposent sur l'hypothèse fondamentale que la règle comptable est respectée. C'est la qualité du système comptable qui est l'objet de l'étude. Un autre champ de littérature s'est développé sur le thème de la « fiabilité » de l'information comptable, c'est-à-dire sur les moyens d'évaluer dans quelle mesure les comptes peuvent être manipulés.

2. La fiabilité du résultat

La problématique de la fiabilité[5] du résultat vient se greffer sur celle de la qualité du résultat comme la mauvaise fée dans le conte de « La Belle au bois dormant ». Et si le résultat publié ne respectait pas les règles ? Ou plus exactement l'esprit des règles ? Face à une telle perspective, toutes les qualités (les dons) données par les gentilles fées (les normes) deviennent alors inopérantes !

Dès lors que les manipulations sont possibles, toute problématique de mesure de la qualité est dépendante de la capacité à déceler et mesurer ces manipulations. Car une information non fiable ne peut pas être de qualité. Autrement dit, il peut être préférable d'avoir une information fiable respectant des règles moins bien adaptées aux besoins de ses utilisateurs, qu'une information manipulée dans un cadre normatif très adapté aux besoins des utilisateurs.

[5] Le terme « fiabilité » opposé à celui de « qualité » est une formulation qui nous est propre. Burgstahler, Hail et Leuz (2006) estiment pour leur part que le terme « qualité comptable » a plusieurs dimensions dont l'une est le conservatisme (en référence au courant « earnings quality »), et qu'une autre dimension est la gestion du résultat. Nous préférons utiliser deux termes différents pour évoquer deux dimensions différentes.

a. Les manipulations comptables

La littérature sur le sujet peine à s'entendre sur une terminologie unifiée et homogène. On trouve le plus couramment l'expression « gestion du résultat » (traduction française de « earnings management ») pour évoquer la capacité qu'ont les dirigeants à orienter les informations comptables publiées en fonction de leurs intérêts ou des intérêts de leur entreprise. La gestion du résultat est ainsi définie par Degeorge, Patel et Zeckhauser (1999) comme étant l'utilisation de la discrétion managériale pour influencer le résultat diffusé auprès des parties prenantes. Mais d'autres expressions sont souvent utilisées pour décrire des concepts ou identiques, ou similaires. L'article de Breton et Stolowy (2004) énumère ainsi et définit des expressions aussi diverses que « manipulations comptables », « transferts comptables de richesse », et des expressions plus imagées comme « window dressing », « big bath accounting », « nettoyage des comptes », « habillage des états financiers » etc.

Le terme de manipulation est souvent interprété comme une fraude, mais la fraude n'est pas l'objet de la thèse. Non pas qu'elle ne soit pas digne d'intérêt, mais elle représente une dimension juridique très spécifique. Par la suite, c'est la vision de Breton et Stolowy (2004) qui est retenue. La manipulation comptable est l'interprétation comptable faite par les dirigeants dans le cadre légal.

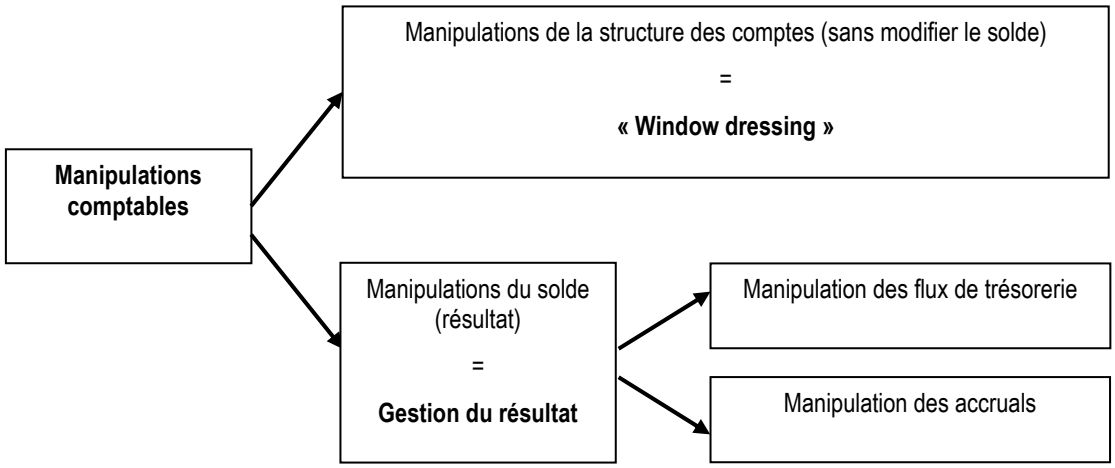
La fiabilité du résultat est donc l'expression réservée à un résultat non manipulé. Un résultat fiable est une condition nécessaire mais pas suffisante à l'obtention d'un résultat de qualité.

La manipulation comptable peut se faire de deux manières (voir Figure 5) :

(1) Soit les dirigeants (responsables de l'élaboration des comptes) jouent sur des règles de présentation des comptes (modifier la structure du bilan - par exemple l'endettement - sans modifier le résultat).

(2) Soit les dirigeants modifient le solde du bilan, c'est-à-dire publient un résultat plus proche de leurs aspirations. On dira alors qu'ils « gèrent le résultat ».

Figure 5 : Les deux types de manipulations comptables



L'expression « gestion du résultat » apparaît dès lors comme un cas particulier de manipulation comptable. Cependant, c'est celui qui est le plus emblématique et généralement le plus lourd de conséquences. C'est pourquoi il focalise la quasi-totalité de la littérature sur le sujet. L'étude des seuils se situe dans ce contexte.

b. La gestion du résultat

La « gestion du résultat » n'est pas la traduction comptable d'un transfert de richesse, mais un choix (une décision) comptable qui a pour conséquence un transfert de richesse. En modifiant l'image de l'entreprise, le dirigeant cherche à influencer des décisions d'allocation de ressources. Ce transfert peut se faire, de l'entreprise vers le propriétaire, de l'entreprise vers le dirigeant, ou de l'entreprise vers un tiers (créancier…) et, réciproquement, des dirigeants, propriétaires ou des tiers vers l'entreprise. La comptabilité n'est plus considérée ici comme un outil neutre de transcription, mais comme un outil d'observation interagissant avec l'objet qu'il est censé observer.

Soit l'équation : Résultat comptable = flux de trésorerie +/- « accruals ». La gestion du résultat peut prendre deux formes.

(1) Elle peut tenter de modifier les flux de trésorerie (anticiper ou retarder des entrées ou sorties). Degeorge, Patel et Zeckhauser (1999) parlent de « direct managing », Glaum, Lichtblau et Lindemann (2004) de « manipulation of cash flows », Roychowdhury (2006) de manipulations des activités réelles, et Burgstahler, Hail et Leuz (2006) de « Business Management ».

(2) Elle peut tenter de jouer sur les options comptables. Les dirigeants agissent sur les accruals. Degeorge, Patel et Zeckhauser (1999) parlent de « misreporting », Burgstahler, Hail et Leuz (2006) de « Reporting Management ».

Depuis les années 80, les chercheurs considèrent généralement qu'il est plus délicat de jouer sur les flux réels. Les études empiriques sur la gestion du résultat ont essentiellement cherché à mettre en œuvre des modèles visant à mesurer les accruals anormaux des entreprises (Healy, 1985 ; DeAngelo, 1986 ; Jones, 1991 ; Dechow, Sloan et Sweeney, 1995 ; Dechow, Richardson et Tuna, 2003). Mais à l'orée des années 2000, un certain nombre d'auteurs ont montré que les activités réelles pouvaient être l'objet de manipulations non négligeables. Des manipulations par accélération des ventes, par retardement des investissements en recherche et développement ou en dépenses de maintenance ont ainsi été identifiées (Fudenberg et Tirole, 1995 ; Healy et Whalen, 1999 ; Dechow et Skinner, 2000 ; Roychowdhury, 2006).

L'enquête menée par Graham, Harvey et Rajgopal (2005) auprès de dirigeants montre, non seulement qu'ils sont prêts à manipuler les activités réelles pour atteindre certains objectifs comme un résultat positif ou les prévisions des analystes, mais qui plus est, qu'ils manifestent une plus grande propension à vouloir manipuler les activités réelles que les accruals. Roychowdhury (2006) souligne que les manipulations via les accruals peuvent difficilement être menées indépendamment de celles sur les activités réelles dans la mesure où, lorsqu'en fin d'exercice, l'écart entre objectif et résultat non manipulé est très important, la manipulation des accruals risque d'être insuffisante pour combler l'écart. Or les manipulations des activités réelles ne peuvent se faire en fin d'exercice. Elles doivent être planifiées à l'avance.

c. La mise en évidence et la mesure de la gestion du résultat

Lorsque les dirigeants cherchent à utiliser leur pouvoir discrétionnaire pour modifier le résultat publié, ils le font en toute « discrétion »[6]. La gestion du résultat est par définition cachée, et il est a priori impossible de la mesurer. Comment étudier (et donc révéler) quelque chose qui est caché ?

[6] La langue française permet ce jeu de mot impossible en anglais.

c.1. Méthodologies statistiques et méthodologies par étude de cas

La méthodologie de l'étude de cas a des limites évidentes. La nature même du phénomène étudié semble incompatible avec une telle démarche, car elle nécessite de chercher au sein de l'entreprise des informations auxquelles le chercheur n'a pas accès. C'est une démarche qui relève davantage d'une procédure d'audit légal ou d'enquête judiciaire.

Certains chercheurs se sont cependant intéressés à des cas de fraude reconnus, soit dans la presse, soit par l'organisme de contrôle, soit par le juge… Mais la méthode de l'étude de cas a d'importantes limites. Elle ne peut être menée qu'a posteriori, lorsqu'un temps suffisamment long s'est écoulé pour que les acteurs acceptent d'en parler. Et même dans ce cas, des réticences et des obstacles importants (techniques, réglementaires et moraux) demeurent. Ainsi, pour obtenir des informations sur d'éventuelles malversations, Prat dit Hauret (2007) interroge des auditeurs à la retraite, des années après les événements. L'étude de cas sur le thème de la gestion du résultat reste donc de portée limitée.

Mais ce qui est caché individuellement peut se révéler sur une grande échelle. C'est la raison pour laquelle les méthodologies statistiques ont été le plus souvent mises en œuvre pour étudier ce phénomène. On peut distinguer deux types de méthodologies statistiques : les enquêtes et les études sur bases de données.

Les études à base de questionnaires peuvent s'avérer extrêmement riches. Mais force est de constater qu'elles sont peu nombreuses sur le sujet de la gestion du résultat. L'étude de Graham, Harvey et Rajgopal (2005) qui interroge directement (et anonymement) des dirigeants sur leurs pratiques, et notamment celles de gestion du résultat, est d'autant plus remarquable qu'elle est unique en son genre.

Les études tentant de faire parler les bases de données sont en définitives de loin les plus courantes sur le sujet. Plutôt que d'étudier les « inputs » qui déterminent l'opacité comptable[7], la recherche se tourne vers les « outcomes » que sont les propriétés des distributions des données comptables publiées à travers les pays et le temps (Bhattacharya, Daouk et Welker, 2003). Ce sont essentiellement ces études qui sont décrites dans la revue de littérature (chapitre II).

[7] Les auteurs utilisent l'expression « earning opacity ».

Une manière tout à fait différente de présenter les études sur la gestion du résultat est de distinguer les études sur des cas avérés (ou suspects), des études sur des entreprises quelconques. Dans le premier cas, on cherche à mesurer comment les entreprises manipulent. Dans le second cas, on cherche à identifier quelles sont les entreprises manipulatrices. Les chercheurs en comptabilité ignorent traditionnellement les approches concernant les cas de fraudes reconnus. Les fraudes reconnues ne recouvrent qu'une partie extrême de la gestion du résultat définie préalablement comme la manipulation effectuée dans le cadre légal. Par ailleurs, les cas avérés ne constituent probablement qu'une faible part des cas de manipulation. Les travaux de Beneish (1999) font figure d'exception. L'auteur propose un modèle prédictif en étudiant des entreprises qui gèrent leur résultat (liste des entreprises identifiées par la SEC ou la presse) et en identifiant des déterminants par un modèle de régression.

Tableau 1 : Typologie des méthodologies d'étude de la gestion du résultat

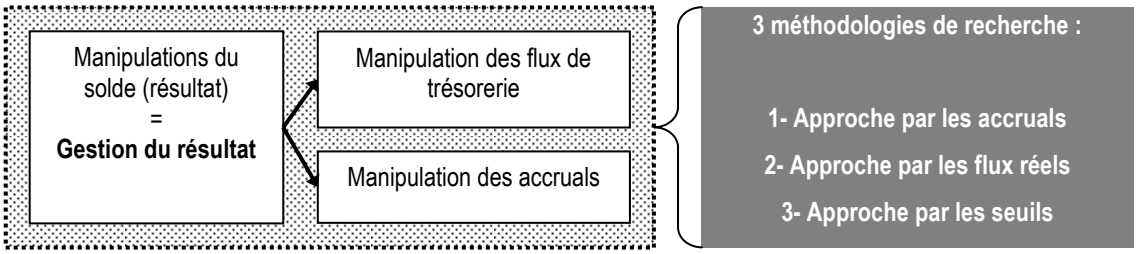
	Méthodologie statistique	**Étude de cas**
Entreprise(s) quelconque(s)	Bases de données : (1a) Mesure des accruals ou des flux de trésorerie irréguliers par des modèles normatifs (1b) Irrégularités statistiques de distribution (seuils) Enquêtes : (2) Questionnaires anonymes sur les pratiques des dirigeants	Étude de cas = impossible par définition ! La gestion du résultat est faite en toute discrétion.
Cas avéré	Les études statistiques sur des cas avérés présentent un intérêt limité	Étude de cas a posteriori

En conclusion, comme le montre le Tableau 1, la plupart des études publiées sur la gestion du résultat sont des études statistiques menées dans l'objectif de révéler des manipulations cachées. Les études de cas, plus rares, semblent plus appropriées aux cas de fraudes avérés.

c.2. Trois méthodologies statistiques

Trois méthodologies statistiques ont été développées par les chercheurs s'intéressant à la gestion du résultat en exploitant des bases de données : les modèles de mesure des accruals, les modèles de mesure des activités réelles, et la mesure des irrégularités de distribution. Dans les trois cas, le chercheur tente d'identifier des indices statistiques qui pourraient lui permettre de penser que les résultats publiés par les entreprises ne sont pas réguliers.

Figure 6 : Les trois démarches statistiques d'étude de la gestion du résultat



Modèles de calcul des accruals anormaux

Dans les modèles d'accruals, le chercheur estime pour chaque entreprise un montant « normal » du résultat.

Soit la relation générale : Résultat comptable = flux de trésorerie +/- « accruals ». Le résultat « normal » peut se décomposer par un montant « normal » des flux réels et un montant « normal » des accruals. Les études sur les accruals postulent que les manipulations de flux réels sont plus difficiles que les manipulations sur les accruals[8]. Les chercheurs focalisent donc leur attention sur cette deuxième composante.

Fondamentalement, les chercheurs ont donc cherché à estimer le montant « normal » des accruals, afin d'en déduire, par soustraction, un montant « anormal » interprété comme la manifestation de manipulations. Les expressions « discrétionnaires » et « non discrétionnaires » sont parfois utilisées à la place de « anormaux » et « normaux » ; ils font référence aux charges et produits qui sont sujets à l'interprétation (la discrétion) des dirigeants et désignent ainsi plus manifestement la nature manipulatrice attribuée aux accruals « anormaux »[9].

Il existe plusieurs modèles d'estimation du montant normal des « accruals ». Les premiers modèles ont estimé les accruals normaux d'une entreprise en référence aux accruals totaux de cette même entreprise durant les années précédentes (Healy, 1985 ; DeAngelo, 1986) ou en référence à un panel d'entreprises de même secteur (Dechow et Sloan, 1991).

Cependant, les modèles qui semblent avoir le meilleur pouvoir explicatif (Dechow, Sloan et Sweeney, 1995) décomposent les accruals en leurs deux dimensions : les variations de

[8] Comme le fait Jeanjean (2002), il faut cependant préciser que les manipulations des flux réels peuvent également des conséquences sur les accruals.

[9] Le terme « anormal » est plus neutre que le terme « discrétionnaire » car ce dernier suppose une intention de la part du dirigeant.

besoin en fonds de roulement, et les autres charges et produits calculés. Ces modèles postulent que les variations de besoin en fonds de roulement doivent évoluer en fonction de l'activité de l'entreprise, alors que les dotations et autres charges calculées doivent évoluer en fonction des investissements. Cette typologie des accruals peut être présentée sous forme de tableau :

Tableau 2 : Typologie des accruals

Accruals	**Normaux** (Non discrétionnaires)	**Anormaux** (Discrétionnaires)
Variations de BFR	Évoluent en fonction du chiffre d'affaires	Manipulations
Dotations (autres charges et produits calculés)	Évolue en fonction des immobilisations	

Une telle décomposition permet de modéliser statistiquement le montant des accruals « normaux » d'une entreprise, et par différence avec le montant réel des accruals, le montant des accruals discrétionnaires (Jones, 1991 ; Dechow, Sloan et Sweeney, 1995 ; Dechow, Richardson et Tuna, 2003).

Enfin, une dernière catégorie d'études fait l'économie d'une modélisation et se contente d'observer l'évolution d'accruals spécifiques comme par exemple le montant des impôts différés (McNichols et Wilson, 1988 ; Beneish, 1997 ; Burgstahler, Elliott et Hanlon, 2003 ; Phillips, Pincus et Rego, 2003 ; Schrand et Wong, 2003 ; Holland et Jackson, 2004 ; Frank et Rego, 2006 ; Petrovits, 2006).

Il est important de souligner les limites méthodologiques des modèles d'estimation des accruals anormaux. La mesure des accruals suppose beaucoup d'hypothèses de calcul et aboutit à des modèles dont le pouvoir explicatif est relativement faible. Par ailleurs, elle ne tient pas pleinement compte de la gestion du résultat qui peut se faire par une manipulation des flux réels (modification des délais de paiement, retard des facturations…), ce qui aboutit à limiter la significativité des résultats obtenus par ces études. Mais si une telle démarche ne permet pas d'avoir une certitude absolue sur la gestion comptable des entreprises, elle obtient des résultats statistiquement intéressants et surtout, permet de mesurer un indicateur de manipulation pour toutes les entreprises de l'échantillon. Autrement dit, la mesure est exhaustive mais peu précise.

Modèle de calcul des activités réelles anormales

Le postulat selon lequel les manipulations des flux réels seraient négligeables a été remis en cause (Fudenberg et Tirole, 1995 ; Healy et Whalen, 1999 ; Dechow et Skinner, 2000). L'enquête de Graham, Harvey et Rajgopal (2005) révèle même que les dirigeants semblent privilégier ce moyen de gestion comptable. Il apparaît donc que les chercheurs ont accordé trop peu d'importance aux modèles de mesure des activités réelles anormales.

Pour répondre à cette lacune, Roychowdhury (2006) développe une méthode empirique pour détecter les manipulations des activités réelles. Il examine les flux de trésorerie opérationnels[10] que sont les coûts de production, et les dépenses discrétionnaires. L'auteur utilise un modèle statistique pour déterminer un montant normal des activités réelles. En comparant le modèle aux charges publiées, il mesure les montants anormaux des activités réelles.

Conceptuellement, cette étude est à rapprocher des modèles sur les accruals (qu'elle complète). Dans les deux cas, les modèles mesurent pour chaque entreprise un niveau « normal » des activités/accruals, et par différence, le montant anormal, donc manipulé. L'approche par les seuils relève, quant à elle, d'une démarche fondamentalement différente.

Approche par les seuils

L'approche par les seuils part d'un postulat : sans manipulation, l'allure de la distribution des résultats ne devrait pas présenter de discontinuité[11]. Autrement dit, le référentiel de qualité de l'ensemble des résultats d'une population est une loi, sinon « Normale », tout au moins « régulière » (c'est-à-dire « lisse »). Ce postulat est formulé pour la première fois par Burgstahler et Dichev (1997).

Contrairement aux deux premières méthodologies qui tentent de mesurer un certain type de manipulation pour chaque entreprise, l'approche par les seuils tente de mettre en évidence des entreprises suspectes, quel que soit le type de manipulation (Activités réelles ou Accruals). Autrement dit, avec les accruals, le chercheur n'est pas tout à fait certain de savoir combien chaque entreprise manipule, alors qu'avec les seuils, le

[10] « Cash Flows from Operations » en anglais.

[11] La discussion sur la nature exacte de la distribution est menée dans la partie 3 de la thèse.

chercheur est à peu près certain de savoir qu'il y a des tricheurs sans savoir exactement qui, ni combien et comment ils manipulent.

L'observation des distributions est ainsi apparue au début des années 2000 comme une alternative à la mesure des accruals pour identifier les pratiques de manipulation, alternative d'autant plus attrayante que l'observation des discontinuités semble a priori facile à effectuer. La partie 2 de la thèse revient sur cet a priori méthodologique.

Enfin, pour Mard (2004) citant Chalayer-Rouchon, Degeorge et Le Nadant (2001), la méthodologie par les seuils est complémentaire à la méthodologie par les accruals car elle « ne permet pas d'apprécier l'ampleur de la gestion des résultats, mais seulement de tester la présence ou non d'un comportement de gestion des résultats ». Les parties 2 et 3 de la thèse reviendront sur cette affirmation.

En conclusion, le chapitre I a situé l'étude des seuils dans le panorama plus général de la recherche en comptabilité. Le chapitre II va dresser un panorama de la littérature sur le sujet.

CHAPITRE II - LITTÉRATURE SUR LES SEUILS

Parmi les études qui se sont intéressées aux seuils comptables, on distingue :

- Celles qui mettent en évidence une irrégularité dans la distribution des résultats publiés par les entreprises et qui interprètent cette irrégularité comme une preuve de manipulation comptable. Elles sont décrites dans la section A.

- Celles qui proposent des explications à ces irrégularités. Elles sont décrites dans la section B.

- Celles qui identifient les moyens mis en œuvre pour manipuler les résultats afin d'éviter un seuil, et leurs conséquences. Elles sont décrites dans la section C.

- Celles qui mesurent l'irrégularité afin d'effectuer des comparaisons dans le temps (évolution), dans l'espace (comparaison internationale), ou entre les différents seuils (hiérarchisation). Elles sont décrites dans la section D.

Enfin, la section E présente les limites conceptuelles et méthodologiques de ces études. Cela permet de préciser le sujet de recherche et justifie son orientation méthodologique.

A. Les travaux observant les seuils

L'approche par les seuils, présentée dans la première sous-partie, est une démarche empirique qui se fonde sur les caractéristiques de la loi de distribution de l'ensemble des résultats. Trois principaux seuils ont été identifiés dans la littérature. Ils sont décrits dans la seconde sous-partie.

1. Une démarche empirique

L'approche par les seuils pose un problème conceptuel : elle se fonde sur une qualité, non pas du résultat individuel d'une entreprise, mais de l'ensemble des résultats d'une population. C'est une approche empirique globale. Le référentiel n'est pas un référentiel du résultat, mais de la distribution d'un grand nombre de résultats.

Le problème, c'est que les règles comptables permettant de définir un résultat sont appliquées de manière individuelle, pour coller à la réalité de chaque entreprise. Elles ne tiennent pas compte des résultats des autres entreprises. Elles en sont totalement indépendantes. Elles ne visent pas à obtenir une loi de distribution homogène quelconque. Il existe des indices de performance construits en relation avec les performances de l'ensemble d'une population. Ce sont des indices de performance relative. C'est le cas du classement Elo aux échecs. C'est le cas du test du quotient intellectuel (QI) ou du TAGE-MAGE (test d'aptitude aux études de gestion de la FNEGE) dans lesquels la moyenne est toujours égale à une certaine valeur. Ces indices visent à mesurer la performance d'un individu par rapport à la performance de ses « concurrents ». Ce n'est pas du tout la logique du résultat comptable.

Conceptuellement, le postulat qui pose comme référentiel de la qualité du résultat le comportement de l'ensemble d'une population suppose que la performance des entreprises obéit à des lois statistiques, loi des grands nombres. Ce postulat pourrait paraître particulièrement fragile si l'observation de la distribution des résultats faite sur des populations différentes, à des époques différentes, ne présentaient pas systématiquement certaines caractéristiques immuables : (1) Une forme « plus ou moins » gaussienne, et (2) des irrégularités (ou discontinuités) autour de certains seuils (toujours les mêmes).

Ces coïncidences empiriques légitiment à elles seules les questions : Comment faut-il interpréter ces caractéristiques ? Peut-on en déduire quelque chose sur la qualité des résultats ? La force de l'approche par les seuils, c'est finalement sa puissance empirique. On constate un phénomène qui est difficilement justifiable théoriquement, mais qui intuitivement, ne semble pas étonnant.

2. Les différents seuils constatés

a. Les travaux précurseurs

Burgstahler et Dichev (1997) sont les premiers à étudier les irrégularités de distribution des résultats comptables. Ils mettent en évidence deux discontinuités sur un échantillon de plus de 4 000 entreprises américaines :

(1) Seuil du résultat nul.

(2) Seuil des variations nulles de résultat.

Dans ces deux situations, les auteurs constatent une sous-représentation du nombre d'entreprises publiant des résultats légèrement en dessous du seuil, et une sur-représentation du nombre d'entreprises publiant des résultats légèrement au dessus du seuil. L'article précurseur de Burgstahler et Dichev (1997) est une référence incontournable pour toutes les études ultérieures traitant des irrégularités de distribution des résultats. Cet article est d'autant plus intéressant que, au-delà de l'observation de l'irrégularité, il apporte une première réponse aux deux questions du pourquoi et du comment. Ces points sont développés plus loin (sections B et C).

Degeorge, Patel et Zeckhauser (1999) ont étudié les irrégularités de distribution de la variable bénéfice par action aux États-unis. Ils confirment les résultats de Burgstahler et Dichev (1997) et identifient un troisième seuil :

(3) Seuil des prévisions des analystes.

Les auteurs constatent que le nombre d'entreprises publiant un résultat légèrement inférieur aux prévisions est anormalement faible alors que le nombre d'entreprises publiant un résultat juste égal, ou légèrement supérieur aux prévisions est anormalement élevé. Cet article est d'autant plus intéressant que, au-delà de l'observation des irrégularités, il apporte une première réponse à la question des conséquences (développée en section C), et propose une hiérarchie des seuils (section D).

Kasznik (1999) a pour sa part mis en évidence un quatrième seuil :

(4) Seuil des prévisions des dirigeants.

Ce seuil est peu étudié en tant que tel. Il est généralement assimilé au troisième seuil (prévisions des analystes). Payne et Robb (2000) montrent par ailleurs que le seuil des

prévisions des analystes est d'autant plus important que leurs prévisions sont en accord avec les prévisions des dirigeants.

Enfin, un cinquième effet est constaté et largement relayé dans la littérature, mais jamais étudié en tant que tel : c'est l'effet « fortes pertes »[12]. Cet effet n'est pas à proprement parler un seuil, dans la mesure où il ne se manifeste pas à un niveau précis de résultat. C'est un effet plus dilué. Il a cependant en commun avec les quatre seuils précédents d'être une irrégularité statistique de distribution et leur est souvent associé dans la littérature. L'effet « fortes pertes » n'est pas l'objet de la thèse, mais il est discuté dans la partie 3 lorsque la nature de la distribution des résultats est abordée.

Dans la thèse, la revue de littérature s'intéresse donc essentiellement aux trois premiers seuils (résultat nul, variation nulle de résultat, erreur nulle de prévision). Ils sont parfois évoqués sous d'autres formulations : évitement des petites pertes, évitement des petites baisses de résultats[13], évitement des petites erreurs de prévision.

Depuis 1999, aucun autre seuil n'a été mis à jour. Cela signifie-t-il qu'il n'en existe pas d'autre ? Ce point est développé dans le chapitre III qui propose une typologie des seuils s'étendant au-delà de ces trois[14] seuils.

b. Les travaux ultérieurs confirment unanimement l'existence des seuils

L'existence de ces trois irrégularités a été largement observée depuis les articles précurseurs, et ce, dans des contextes variés. Si la plupart des études concernent les entreprises américaines, les seuils ont été observés également auprès d'entreprises françaises cotées entre 1990 et 1998 (Mard, 2004), auprès d'entreprises chinoises (Yu, Du et Sun, 2006), allemandes (Glaum, Lichtblau et Lindemann, 2004), russes (Goncharov et Zimmermann, 2006), australiennes (Holland et Ramsey, 2003 ; Davidson, Goodwin-Stewart et Kent, 2005), et japonaises (Shuto, 2008).

[12] Ou « big bath accounting », ou « grande lessive »… Les dirigeants d'une entreprise, lorsqu'ils sont obligés à annoncer une mauvaise nouvelle, peuvent en profiter pour noircir exagérément la situation de manière à se dégager des marges de manœuvre pour le futur.

[13] Sans se confondre totalement avec, cet évitement est à rapprocher du lissage à la hausse des résultats.

[14] Par la suite, le seuil des prévisions des dirigeants sera assimilé au troisième seuil (prévisions des analystes) pour ne retenir que les trois principaux seuils.

La plupart des études s'intéressent aux entreprises cotées, mais les seuils ont également été étudiés auprès d'entreprises non cotées françaises (Bisson, Dumontier et Janin, 2004), et européennes (Coppens et Peek, 2005 ; Burgstahler, Hail et Leuz, 2006).

Par ailleurs, les irrégularités sont apparentes dans tous les secteurs d'activité. Leur existence a été observée auprès d'échantillons composés exclusivement de banques, secteur généralement exclu des études empiriques sur bases de données (Schrand et Wong, 2003 ; Shen et Chih, 2005).

En conclusion, tous les articles publiés sur le sujet observent des irrégularités de distribution autour des trois seuils. Aucun article n'a jamais remis en cause leur existence. Les études sur les seuils ne se limitent cependant jamais à la simple constatation des seuils. Ils vont aborder la question des causes de ces seuils (section B), identifier les moyens mis en œuvre pour éviter les seuils, et leurs conséquences (section C), ou mesurer les irrégularités afin de les comparer (section D). Ces différents travaux sont présentés et leur synthèse (section E) permet de préciser la question de recherche étudiée dans la thèse.

B. Les travaux sur les causes des seuils

Les irrégularités de distribution de résultat sont interprétées comme étant la manifestation de manipulations comptables. Les articles précurseurs ont proposé des explications, qui sont depuis reprises telles quelles par l'ensemble des chercheurs. La première sous-partie présente le postulat de manipulation ; la deuxième présente les causes évoquées pour expliquer ces manipulations.

1. L'irrégularité est la manifestation d'une manipulation

a. Le seuil du résultat nul : un postulat de manipulation

Hayn (1995) est le premier auteur à identifier comme suspect le nombre insuffisant d'entreprises faiblement déficitaires. À la différence de Burgstahler et Dichev (1997),

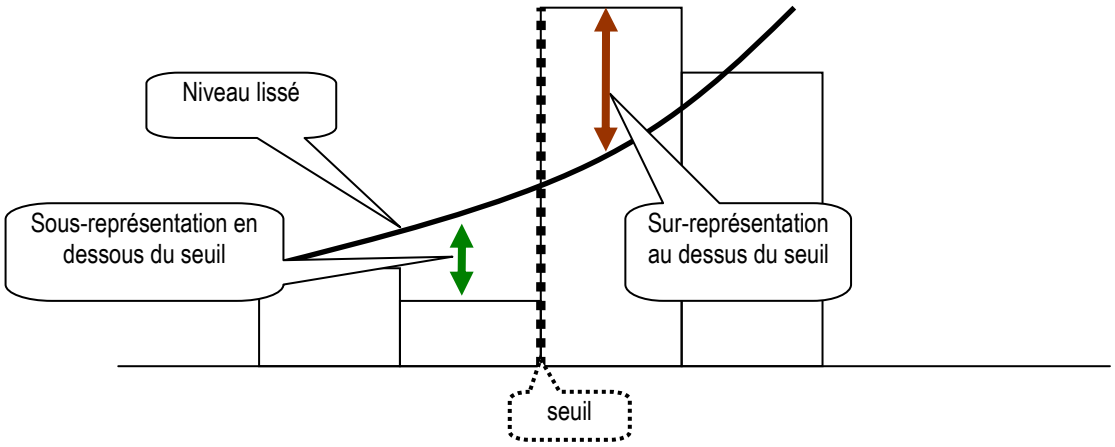
l'auteur n'étudie pas la distribution dans son ensemble, et n'observe pas le surnombre à droite du seuil. Elle en déduit cependant que si le nombre d'entreprises faiblement déficitaire est anormalement faible, c'est que les entreprises évitent de publier un résultat négatif. Dépasser le seuil du résultat nul correspond à l'envoi d'un signal positif vers le marché, et rater le seuil correspond à l'envoi d'un signal négatif.

Cette présomption de manipulation est d'autant plus généralement admise qu'elle est reconnue par les dirigeants eux-mêmes. C'est ce que révèle notamment l'enquête menée par Graham, Harvey et Rajgopal (2005).

b. La manipulation est-elle marginale ?

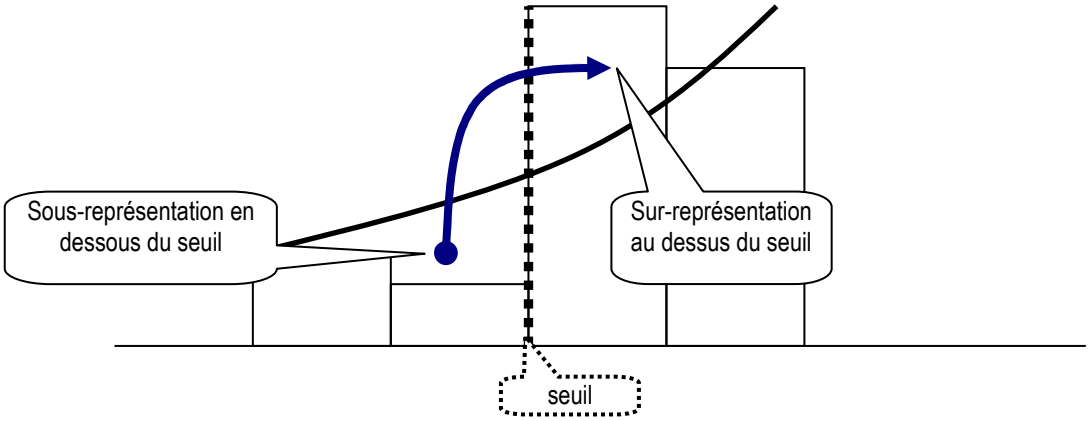
L'étude de Burgstahler et Dichev (1997) étend les observations de Hayn (1995) à l'ensemble de la distribution des résultats, et à l'étude de la distribution des variations de résultat des entreprises américaines. Ils observent alors que parallèlement à l'absence d'entreprises faiblement déficitaires, le nombre d'entreprises faiblement bénéficiaires est, lui, étonnamment élevé.

Figure 7 : La double discontinuité



Les auteurs proposent alors l'interprétation suivante : puisqu'il y a une sous-représentation des effectifs juste en dessous du seuil, et une sur-représentation des effectifs juste au dessus, c'est qu'il y a un transfert d'effectifs d'un intervalle sur un autre. Les entreprises calculant un résultat (non manipulé) juste inférieur au seuil sont très motivées pour manipuler leur résultat afin de publier un résultat légèrement supérieur au seuil. Autrement dit, les irrégularités mesurent des petites gestions de résultat à la hausse.

Figure 8 : Interprétation de la double discontinuité



Un exemple tiré de la presse économique illustre cette interprétation. Dans le Wall Street Journal, McGough (2000) cite le cas de la société Black Box[15] qui publie en 2000 un résultat par action de 72 cents, exactement en accord avec les prévisions des analystes. En réalité, le résultat précis est de 71,505 cents par action, et c'est grâce à un arrondi que le chiffre publié correspond aux prévisions. Cela signifie qu'au niveau du résultat global, à 750,62 $ près, le seuil de prévision des analystes n'aurait pas été atteint. Cet exemple est cité par Das et Zhang (2003) pour démontrer que le montant de la manipulation pour éviter un seuil peut être très faible.

Cependant, si cet exemple illustre concrètement le fait que le montant à manipuler n'est pas nécessairement élevé, il ne permet pas de démontrer pour autant, ni qu'il y a eu manipulation, ni que, si manipulation il y a eu, le montant manipulé ait été exactement de 750,62 $. Il est tout à fait possible que le résultat avant manipulation ait été très inférieur à 71 cents par action. Il est donc important de souligner que l'interprétation de Burgstahler et Dichev (1997) n'est pas démontrée. C'est un postulat.

Malgré une certaine diversité, presque toutes les études reprennent le postulat de manipulation marginale. Plus précisément, la littérature sur les seuils présente trois catégories d'articles :

(1) Les articles les plus nombreux évoquent explicitement le postulat de faible manipulation (Burgstahler et Eames, 2003 ; Das et Zhang, 2003 ; Dechow, Richardson et Tuna, 2003 ; Leuz, Nanda et Wysocki, 2003 ; McNichols, 2003 ; Coppens et Peek, 2005 ; Beaver, McNichols et Nelson, 2007). Ces articles parlent en général de gestion des résultats pour éviter les « faibles » pertes[16]. Ils reprennent à leur compte l'explication

[15] Le nom de la société était-il prédestiné ?

[16] En anglais : « Earnings manipulations to avoid small losses ».

fournie par Burgstahler et Dichev (1997) dans leur article fondateur : « Nos résultats confirment l'hypothèse de manipulation des flux de trésorerie d'exploitation afin de transformer les faibles pertes en résultats faiblement positifs[17] ». Ces explications sont également présentes dans la littérature francophone (Mard, 2004). Ainsi, Jeanjean (2001) écrit : « Il en résulte une aversion pour les pertes d'un montant faible (…) ».

(2) Quelques articles, peu nombreux, ne posent aucune hypothèse sur les montants manipulés. Ces articles parlent généralement de gestion des résultats pour éviter les pertes[18]. C'est notamment le cas de l'article de Degeorge, Patel et Zeckhauser (1999). Cet article est cependant souvent cité à tort. Ainsi, Leuz, Nanda & Wysocki (2003) écrivent : « Degeorge et al. (1999) […] montrent que les dirigeants d'entreprises américaines utilisent leur discrétion en matière de choix comptables afin d'éviter de publier des petites pertes[19] ». L'expression « petites pertes » ne figure pas dans l'article cité. De tels propos sont également présents dans les articles de Coppens et Peek (2005), et de McNichols (2003) ainsi que dans un grand nombre d'articles plus récents qui s'inspirent des précédents.

(3) Enfin, certains articles ne font pas explicitement référence au montant manipulé, mais postulent implicitement que la manipulation pour atteindre le résultat positif est de faible ampleur. Peuvent être notamment classés dans cette catégorie tous les articles qui se fondent sur un indicateur d'asymétrie (Leuz, Nanda et Wysocki, 2003 ; Glaum, Lichtblau et Lindemann, 2004 ; Brown et Caylor, 2005 ; Burgstahler, Hail et Leuz, 2006). En effet, ces études mesurent l'ampleur de la manipulation comptable en faisant un rapport entre l'effectif immédiatement à droite du seuil sur l'effectif immédiatement à gauche du seuil. Autrement dit, elles assimilent la sous-représentation à gauche et la sur-représentation à droite à l'expression du même phénomène. C'est donc un postulat implicite de manipulation de faible ampleur du résultat.

En conclusion, à l'exception notable de Degeorge, Patel et Zeckhauser (1999), la littérature comptable tend quasi unanimement à postuler que la gestion du résultat pour éviter les pertes, et de manière plus générale la gestion du résultat pour atteindre des seuils, s'effectue sur de petits montants. C'est le postulat de « manipulation marginale » pour atteindre des seuils. Le corollaire de ce postulat est que les manipulations qui

[17] Notre traduction de : « We find evidence consistent with the manipulation of cash flow from operations to effectively move observations from small losses to small positive earnings ».

[18] En anglais : « Earnings manipulations to avoid losses ».

[19] Notre traduction de : « Degeorge et al. (1999) […] present evidence that U.S. managers use accounting discretion to avoid reporting small losses ».

permettent aux entreprises d'éviter un seuil comptable sont d'autant plus difficiles à identifier qu'elles sont de petite ampleur.

Si le postulat de manipulation pour atteindre ou dépasser un seuil est reconnu par les dirigeants (Graham, Harvey et Rajgopal, 2005), les montants évoqués ne sont pas nécessairement faibles. Le postulat de manipulation marginale est donc sans doute une évidence admise trop rapidement.

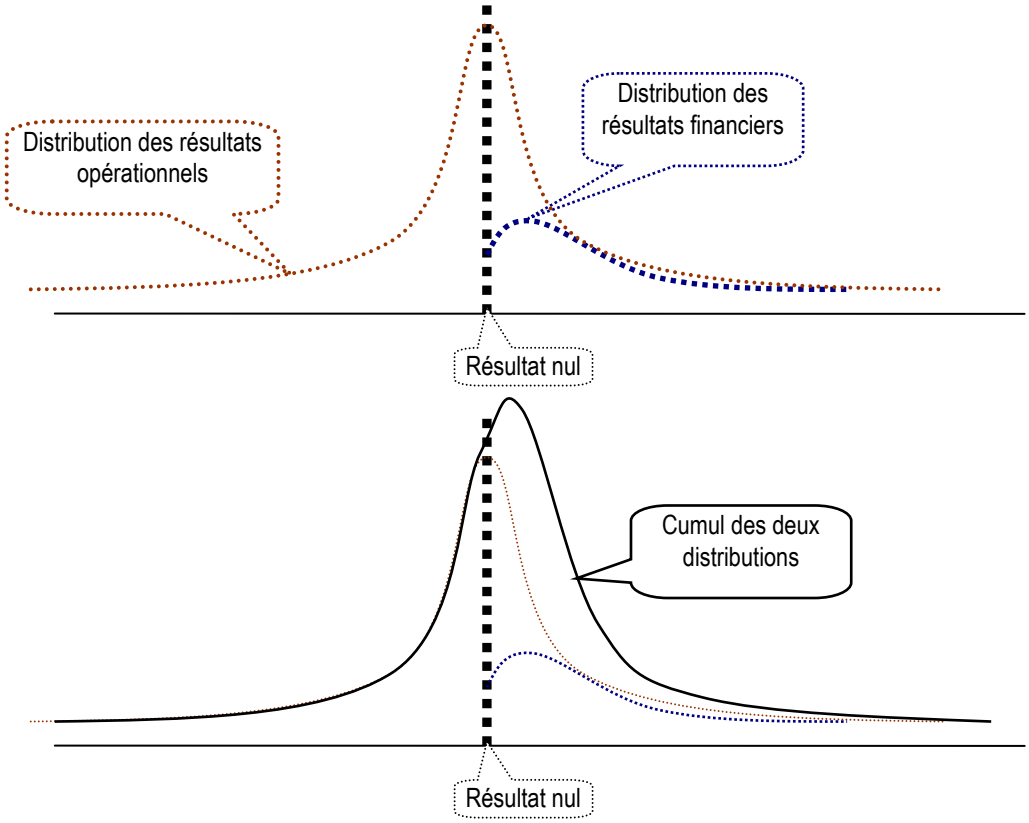
c. Des explications alternatives

Si aucun auteur n'a pu démontrer l'inexactitude du postulat de manipulation, certains ont relativisé sa portée en proposant des explications, sinon alternatives, du moins complémentaires. Dechow, Richardson et Tuna (2003) identifient ainsi 5 pistes de réflexion.

c.1. Les actifs financiers

Dechow, Richardson et Tuna (2003) proposent une piste financière à l'explication du seuil résultat nul. D'après eux, l'utilisation d'actifs financiers pourrait affecter la distribution des résultats. En effet, les actifs financiers génèrent des intérêts ou des dividendes qui ne peuvent être négatifs. Autrement dit, la distribution des seuls résultats financiers est tronquée au zéro.

Figure 9 : Les actifs financiers comme explication à l'irrégularité



La distribution des résultats peut alors être vue comme la somme d'une distribution de résultats opérationnels (non irrégulière au zéro) et d'une distribution de résultats financiers (tronquée au zéro) comme l'illustre le second graphique de la Figure 9. Cela conduit alors à la sur-représentation des effectifs des intervalles juste supérieurs à zéro. Cette explication n'a pas été testée.

c.2. La théorie de l'effort

Dechow, Richardson et Tuna (2003) évoquent la « théorie de l'effort » : Une entreprise qui sait que son résultat sera négatif va tenter de mobiliser les énergies afin d'éviter cette situation. Ce surplus de motivation des salariés dans les périodes « critiques » ou « exceptionnelles » est une explication alternative à la gestion du résultat puisque l'irrégularité constatée ne serait pas due à des manipulations, mais (en partie ou totalement) à la traduction comptable sincère d'une réelle performance économique.

Cependant, il est possible d'objecter que les entreprises qui demandent un effort particulier à leurs salariés en période difficile le font avant de connaître avec exactitude leur résultat. Autrement dit, quand bien même il y aurait un effet « effort » ponctuel en

période critique, il n'est pas le fait unique des entreprises faiblement déficitaires, mais d'un ensemble d'entreprises dont le risque de contre-performance est jugé important par les dirigeants, et ce, sans connaître à l'avance le résultat de l'exercice. En l'occurrence, si un effet « effort » doit se faire sentir, il peut justifier la dissymétrie et la concentration autour de la moyenne, mais pas la discontinuité qui se manifeste exactement autour du résultat nul.

c.3. Les contraintes réglementaires

Les entreprises cotées doivent remplir des conditions et notamment des conditions de profitabilité pour entrer dans la cotation. Dechow, Richardson et Tuna (2003) soulignent donc le risque d'un biais de sélection des entreprises dans les études réalisées généralement auprès d'entreprises cotées. Les auteurs montrent (graphiquement uniquement) que le pic semble plus grand pour les entreprises cotées depuis moins de deux ans que pour les entreprises cotées depuis plus de 20 ans.

Cette limite ne remet pas en cause le postulat de manipulation pour éviter les seuils. Elle en souligne seulement une des causes possible.

c.4. La prudence des règles comptables

La prudence[20] des règles comptables favoriserait la publication de résultats faiblement positifs. En effet, la prudence a tendance à décaler vers la gauche de la courbe les entreprises qui détiennent un actif qui ne génère pas de rentrées (en dotant des provisions) alors qu'elle tend à sur-représenter les entreprises qui détiendraient le même actif profitable juste au dessus du zéro (en étalant la charge dans le temps). Autrement dit, le principe de prudence expliquerait le caractère concentré des distributions et la présence des queues épaisses de distribution.

Si la prudence des règles comptables a été souvent évoquée pour expliquer la dissymétrie des distributions de résultats (Givoly et Hayn, 2002[21] ; Bhattacharya, Daouk et

[20] « Conservatism » en anglais.

[21] Dans un article sur le conservatisme dans les états financiers, les auteurs s'interrogent sur l'interprétation à donner à la dissymétrie de la dispersion des résultats. D'après les auteurs, il n'y a pas de définition du conservatisme qui fasse autorité. Mais la notion de conservatisme suggère que les mauvaises nouvelles (pertes ou risques) sont prises en compte (traduites dans les états financiers) plus rapidement que les bonnes nouvelles (produits). C'est une autre manière de nommer le principe de prudence. Ainsi, la forte queue de distribution traduit la méfiance (comptable) vis-à-vis d'entreprises à risque, alors que le faible étalement à droite traduit la résistance à traduire comptablement les bonnes performances. Les auteurs observent aux États Unis, durant les 20 dernières années, une tendance au renforcement du conservatisme des états financiers. Pour effectuer leurs observations, les auteurs utilisent plusieurs mesures. L'une d'elles est le « Skewness of earnings » qui est une mesure de la dissymétrie de la dispersion. Son

Welker, 2003), son évocation pour justifier les irrégularités autour du résultat nul est plus délicate. En effet, en retardant la prise en compte des bonnes nouvelles, et en anticipant les mauvaises nouvelles, la prudence permet de justifier la présence de fortes pertes, ou la concentration autour de la moyenne. Mais pourquoi le résultat nul serait-il le pivot statistique de la prudence ?

c.5. L'influence de la variable de mise à l'échelle

Toutes les études observant les distributions de résultat mettent ces résultats à l'échelle par une variable de taille. Or les irrégularités observées sur la distribution de la variable mise à l'échelle peuvent être causées, non pas par le résultat (numérateur) mais par la variable de mise à l'échelle (dénominateur). Parmi ces dénominateurs, de nombreuses études utilisent la valeur boursière en début d'année de l'entreprise. L'évaluation des entreprises faiblement déficitaires n'est pas nécessairement faite selon la même méthode que pour les entreprises faiblement bénéficiaires (Dechow, Richardson et Tuna, 2003 ; Durtschi et Easton, 2005). Par exemple, les évaluations pourraient être plus orientées sur les valeurs liquidatives du bilan. Les auteurs en déduisent que mettre à l'échelle par la valeur de marché doit accentuer l'irrégularité.

Cette explication, sans remettre en cause l'interprétation de manipulation, attire l'attention sur une limite méthodologique importante des études sur les seuils. Elle sera développée dans la partie 2 de la thèse.

d. Le postulat est étendu aux autres seuils (prévisions et variations)

De même que l'irrégularité autour du résultat nul est interprétée comme la manifestation d'une gestion du résultat, l'irrégularité autour des deux autres seuils (évitement d'une petite baisse de résultat et évitement d'une petite erreur de prévision) est interprétée de la même manière. Les entreprises manipulent leurs comptes (a priori de manière marginale) afin d'atteindre et dépasser ces seuils.

Cependant, si les deux premiers seuils font référence à un niveau objectif de résultat (résultat = zéro et résultat = résultat de l'année précédente), le seuil des prévisions fait, quant à lui, référence à un niveau subjectif. Les prévisions sont des opinions formulées par des analystes (ou les dirigeants eux-mêmes). Ces prévisions peuvent être

évolution tend à montrer qu'il était légèrement positif dans les années 50 à 70, et que depuis les années 80, il a tendance à devenir de plus en plus fortement négatif. Cette dissymétrie est interprétée comme un signe de la croissance du conservatisme des états financiers.

divergentes, et sont susceptibles d'évoluer. Dans l'équation : erreurs de prévisions = résultats prévus – résultats publiés, si les erreurs nulles sont anormalement nombreuses, c'est que (1) soit les résultats publiés sont manipulés, (2) soit les résultats prévus sont orientés. Autrement dit, les dirigeants peuvent aiguiller les analystes vers des objectifs de résultat que la société est en mesure d'atteindre ou de dépasser.

Un certain nombre d'études a ainsi cherché à déterminer dans quelle mesure l'irrégularité devait être attribuée à une gestion du résultat, ou à une gestion des prévisions[22]. Les auteurs montrent empiriquement que les analystes sont manipulés/influencés/guidés par les dirigeants (Bartov, Givoly et Hayn, 2002 ; Matsumoto, 2002 ; Burgstahler et Eames, 2006[23] ; Cotter, Tuna et Wysocki, 2006). Sur ce sujet, il est judicieux de remarquer que, si les dirigeants peuvent être motivés à éviter de publier des résultats supérieurs ou égaux aux prévisions, les analystes ne sont pas moins motivés à éviter de faire des erreurs. Une prévision juste est considérée comme une prévision de qualité et la réputation des analystes est d'autant plus forte que leurs prévisions sont vérifiées (Chambost, 2007).

Enfin, Wasley et Wu (2005) montrent que les entreprises qui ratent le seuil des prévisions tentent d'atténuer l'impact d'une telle mauvaise nouvelle en publiant des prévisions de flux de trésorerie favorables. La manipulation du résultat n'apparaît donc pas comme le seul recours pour éviter l'impact négatif d'un ratage de seuil sur le marché.

2. Les études des causes des irrégularités

Les articles fondateurs font allusion à cette question et proposent une base théorique explicative. Mais il est notable que la littérature n'a, depuis, pas développé ce cadre explicatif. Les chercheurs se contentent de citer les articles précurseurs de Burgstahler et Dichev (1997) et Degeorge, Patel et Zeckhauser (1999). Ils délaissent la question du pourquoi et tiennent pour acquis qu'il y a irrégularité et que l'irrégularité est signe de manipulation.

[22] Dans la littérature anglophone, on trouve les expressions « forecast management », « forecast guidance », ou encore « expectations management ».

[23] Dans un article précédent, Burgstahler et Eames (2003) estimaient que les analystes anticipaient la gestion du résultat destinée à éviter les faibles pertes et les faibles variations négatives. Cette « anticipation » s'est transformée en présomption de « guidance » avec le temps.

Précisément, que cherche-t-on à expliquer ? La question « pourquoi les entreprises manipulent-elles leurs comptes si le résultat (non manipulé) est inférieur à un objectif ? » peut se décomposer en deux sous-questions : pourquoi… (1) …manipule-t-on le résultat... (2) …pour atteindre un seuil ? La première partie de la question renvoie à l'explication généralement apportée à la gestion du résultat, alors que la seconde partie renvoie à des explications spécifiques aux seuils. Le postulat sous-jacent est que toutes les pratiques de gestion du résultat ne visent pas nécessairement à atteindre un seuil.

Ce double questionnement permet de construire le plan de la section. Les causes de la gestion du résultat sont abordées dans la sous-partie a, puis des explications spécifiques aux seuils sont présentées dans les sous-parties b (théorie psychologique et théorie prospective) et c (théorie des coûts de transaction). La sous-partie d présente l'arbitrage coût/avantage et la dernière sous-partie (e) présente les travaux qui identifient des facteurs environnementaux influençant les pratiques d'évitement des seuils.

a. Le cadre théorique de la recherche positive sur la gestion du résultat

Les méthodologies essentiellement statistiques mises en œuvre pour étudier les seuils ont naturellement intéressé les chercheurs en comptabilité les plus rodés avec l'outil statistique[24]. Ces travaux s'inscrivent généralement dans la théorie dite positive de la comptabilité, cadre généralement mobilisé par les travaux statistiques à partir de bases de données en comptabilité.

Ce cadre n'est cependant jamais explicitement évoqué dans les articles sur les seuils. Il apparaît cependant qu'il propose des explications à la gestion du résultat pour atteindre et dépasser les seuils. Ce cadre appelé « théorie politico-contractuelle de la comptabilité » (Raffournier, 1990) est désormais très standardisé. Proposé par Watts et Zimmerman (1978), il cherche à expliquer les décisions comptables à partir des relations d'agence et des coûts politiques auxquels les entreprises sont soumises (Dumontier et Raffournier, 1999).

Les coûts d'agence

Dans le cadre de la théorie des contrats et de la théorie de l'agence de Jensen et Meckling (1976), l'opportunisme des agents les conduit à optimiser leur utilité personnelle

[24] Quand bien même les méthodologies ne reposent pas sur des régressions.

sous contrainte de rationalité limitée. La coordination des actions individuelles s'effectue par la conclusion de contrats, formels ou informels, définissant les obligations de chacun et les rémunérations correspondantes. Les dirigeants qui ne possèdent pas la totalité des actions de leur entreprise ont intérêt à consommer une partie des richesses de la firme sous forme de satisfaction personnelle (rémunération personnelle, avantages en nature…). Pour réduire cette incitation, la rémunération des dirigeants est souvent indexée sur le résultat. Il en découle la première proposition (Watts et Zimmerman, 1986) : Les entreprises dans lesquelles la rémunération des dirigeants comporte une clause d'intéressement aux résultats choisissent les méthodes comptables qui accélèrent la comptabilisation des bénéfices. Dans le cadre de l'étude des seuils, il apparaît que dès qu'une clause contractuelle est exprimée en fonction d'un niveau de résultat (prime accordée aux dirigeants si le résultat atteint un certain niveau…), ce niveau peut constituer un seuil de publication.

La théorie de l'agence est plus spécifiquement invoquée par Kasznik (1999) pour expliquer l'évitement des surprises négatives. L'auteur estime que, pour un dirigeant, ne pas atteindre ses prévisions est une source potentielle de coûts : actions en justice de la part des investisseurs, ou perte de réputation. Les dirigeants gèrent donc leur résultat à la hausse lorsque les résultats (non gérés) sont inférieurs aux prévisions. Si cette explication s'intègre dans le cadre des motivations contractuelles, elle l'élargit en soulignant que la motivation des dirigeants n'est pas uniquement liée à leur rémunération, mais peut être également liée à leur notoriété.

Les actionnaires ne sont pas les seuls investisseurs de l'entreprise. Les créanciers sont eux aussi partie prenante des relations contractuelles. Or l'ensemble des ressources de l'entreprise est géré par les actionnaires par l'intermédiaire des dirigeants qu'ils désignent. Les actionnaires peuvent donc détourner à leur profit une partie des fonds apportés par les créanciers, en leur faisant supporter une partie des risques liés à leurs décisions d'investissement. Les créanciers se protègent en incluant dans les contrats d'endettement des clauses limitatives[25] sur les versements de dividendes ou les ratios d'endettement. Il en découle la seconde proposition (Watts et Zimmerman, 1986) : Les entreprises les plus proches des limites imposées par les contrats de prêt choisissent des méthodes qui diminuent la probabilité d'une violation de ces clauses. Cette proposition justifie clairement l'existence de certains seuils (ou objectifs), même si les seuils identifiés jusqu'alors par la littérature (résultat nul, variation nulle de résultat et erreur nulle de prévision) ne semblent pas être expliqués par l'existence des clauses d'endettement.

[25] L'expression « clauses restrictives » est également utilisée. En anglais, l'expression utilisée est « Covenants ».

Autrement dit, il est intéressant de souligner que la théorie prévoit l'existence de seuils non encore observés empiriquement.

Les coûts politiques

Dans le cadre de la théorie des coûts politiques (Watts et Zimmerman, 1978), baptisée théorie de la réglementation par Healy et Whalen (1999), le processus politique est une compétition pour le partage des richesses. Les entreprises qui produisent de la richesse sont soumises à des ponctions d'autant plus importantes que les bénéfices sont élevés. En effet, un résultat anormalement élevé peut être interprété comme la manifestation d'une rente monopolistique ou oligopolistique qui justifie un prélèvement complémentaire ou l'instauration de nouvelles contraintes réglementaires. Certaines entreprises, de par leur taille ou de par leur secteur d'activité sensible, sont plus surveillées que d'autres, et donc plus sujettes à des coûts politiques. Il en découle la troisième proposition de la théorie politico-contractuelle : Les entreprises les plus sujettes à des coûts politiques choisissent les méthodes comptables qui retardent la comptabilisation des bénéfices. Par de telles manipulations, on dit que les entreprises gèrent leurs coûts politiques. Dès lors que les réglementations sectorielles, ou anti-trust s'expriment en référence à un certain niveau de résultat, ces niveaux constituent des seuils de publication.

À l'exception de Yu, Du et Sun (2006) qui mettent en évidence une irrégularité autour d'une rentabilité minimum réglementaire que les entreprises chinoises cotées doivent respecter, les seuils de publication mis en évidence jusqu'alors par la littérature n'ont pas été directement expliqués par les coûts politiques. La théorie prévoit donc à nouveau l'existence de seuils pas (ou peu) observés empiriquement.

Il est possible d'associer à ces motivations réglementaires les motivations fiscales qui peuvent inciter les entreprises, toutes choses égales par ailleurs, à reporter dans le temps leurs bénéfices. Plus concrètement, dans le cadre de la gestion des résultats pour éviter les seuils, il apparaît que dans la plupart des cas, un résultat négatif (ou plus précisément un résultat fiscal négatif) permet d'éviter de payer de l'impôt. Une telle motivation peut justifier une surreprésentation des entreprises faiblement déficitaires, sous réserve d'observer la distribution des résultats fiscaux (et en tenant compte des éventuels reports déficitaires). Ce qui peut s'exprimer sous forme de proposition : Les entreprises dont une variation du résultat comptable génère une variation de l'impôt choisissent les méthodes comptables qui retardent la comptabilisation des bénéfices. Une telle explication est évoquée dans le cadre de l'observation des résultats d'entreprises de petite taille, ne

faisant pas appel public à l'épargne (Bisson, Dumontier et Janin, 2004 ; Coppens et Peek, 2005).

Mais la difficulté d'observation est liée au fait que dans bien des pays, le résultat comptable ne sert pas directement au calcul de l'impôt. Le résultat comptable diffère du résultat fiscal. Seule l'étude de Goncharov et Zimmermann (2006) menée auprès d'un échantillon d'entreprises russes avant et après une loi rendant la comptabilité indépendante des règles fiscales en 2001 a été menée sur le sujet. Les auteurs confirment que la politique fiscale peut avoir une influence sur la gestion du résultat pour éviter le seuil du résultat nul.

En définitive, la théorie positive permet d'apporter certaines explications aux pratiques de manipulation pour éviter certains seuils (objectif de rémunération, seuil réglementaire, fiscal, ou induit par des clauses d'endettement). Mais elle n'est jamais évoquée car elle n'est pas adaptée pour expliquer les trois principaux seuils observés (résultat nul, variation nulle de résultat, erreur nulle de prévision). C'est pourquoi la recherche comptable sur les seuils, bien que faisant partie du courant de recherche positive en comptabilité, s'est développée sans faire référence directement au cadre fondateur de ce même courant. Les auteurs ont donc proposé des explications plus spécifiques fondées sur (1) le comportement et la psychologie des agents économiques, et sur (2) les coûts générés par un résultat publié inférieur au seuil (ou « ratage » d'un seuil).

b. Les facteurs spécifiques comportementaux

Dans leur article fondateur, Burgstahler et Dichev (1997) proposent deux cadres théoriques pour analyser les motivations des dirigeants qui cherchent à éviter les seuils identifiés. Le premier fait appel à la théorie prospective[26]. Il est décrit ci-dessous sous l'intitulé plus large des « facteurs comportementaux » car l'argument de la psychologie des chiffres négatifs cité par Degeorge, Patel et Zeckhauser (1999) lui est associé. Le second cadre théorique, fondé sur les coûts de transaction est développé dans la sous-partie suivante (c.).

La théorie prospective de Kahneman et Tversky (1979) suggère que les acteurs prennent leurs décisions en faisant toujours inconsciemment référence à un état. Aux alentours de cet état, des décisions binaires seront prises de manière non « progressive ». C'est le cas notamment lorsque l'on passe d'une situation de perte relative ou absolue à un gain. De

[26] Cette théorie est généralement connue sous son intitulé anglais de « Prospect Theory ».

par sa valeur emblématique, un résultat nul sera plus souvent considéré par les investisseurs comme un état de référence que toute autre valeur de résultat. Il en va de même de la prévision de résultat ou du résultat de N-1 qui peuvent servir d'état de référence. Un nombre important d'acteurs vont donc se forger une opinion et prendre leurs décisions concernant une entreprise sur le critère dichotomique : le seuil a-t-il été franchi ou non ? Cette théorie explique le rôle clef que peuvent jouer certains seuils de publication comptable (notamment auprès des marchés financiers), mais elle ne suffit pas à expliquer le mécanisme par lequel ces seuils se traduisent par des manipulations comptables.

Degeorge, Patel et Zeckhauser (1999) mettent l'accent sur les motivations des dirigeants en invoquant la valeur symbolique et psychologique des chiffres négatifs mise en évidence (Cornsweet, 1974 ; Glass et Holyoak, 1986). Les auteurs se servent notamment de ce cadre pour justifier la hiérarchie entre les trois seuils constatés. Les entreprises privilégient dans l'ordre décroissant un résultat faiblement positif, puis un résultat en croissance, puis le respect des prévisions. Cette hiérarchisation des priorités est un indicateur intéressant de compréhension des motivations des dirigeants. Cependant, cette hiérarchie peut évoluer dans le temps, influencée par le rôle des médias ou de la réglementation boursière (Dechow, Richardson et Tuna, 2003 ; Brown et Caylor, 2005). La valeur psychologique des chiffres négatifs ne suffit donc pas à elle seule à expliquer la hiérarchie des seuils.

De manière plus générale, la notion de prix psychologique n'est pas étrangère à la politique de publication des comptes. Carslaw (1988) a montré auprès d'un échantillon d'entreprises néo-zélandaises que les dirigeants publient des résultats se terminant plus rarement par des 9 et plus souvent par des 0 lorsqu'ils sont positifs, et l'inverse lorsqu'ils sont négatifs. Ainsi, une perte de 39,95 € est perçue comme nettement moins importante qu'une perte de 40 € et un bénéfice de 40 € est perçu comme nettement plus important qu'un bénéfice de 39,95 €. Ce constat a été étendu par Thomas (1989) aux entreprises américaines.

Dans tous les cas, que ce soit pour leur valeur symbolique, ou pour leur statut de référence, certains niveaux de résultat semblent pouvoir jouer un rôle de seuil décisionnel pour les parties prenantes de l'entreprise. Ces explications comportementales ne suffisent cependant pas à elles seules à justifier les pratiques de gestion du résultat. Elles doivent être combinées à la présence de coûts spécifiques à la proximité des seuils pour prendre leur véritable pouvoir explicatif.

c. Les coûts de transaction

Dans leur article fondateur, Burgstahler et Dichev (1997) évoquent le coût des transactions de l'entreprise avec ses parties prenantes[27] pour expliquer les manipulations comptables pour atteindre les seuils. Cette explication est adoptée et développée par Glaum, Lichtblau et Lindemann (2004). La plupart des articles ultérieurs ne remettent pas en cause cette explication. Ils se contentent de citer les articles précurseurs reprenant ainsi à leur compte leurs explications.

Formulée par ces auteurs, la théorie des coûts de transactions se fonde sur deux arguments : (1) Il existe un coût d'obtention des informations pour les parties prenantes qui a pour conséquence qu'un nombre suffisamment important parmi eux se contente de l'information publiée sur le résultat pour prendre des décisions. (2) Le résultat comptable publié affecte les relations entre la firme et les parties prenantes. En général, un résultat élevé permet de négocier des transactions dans des conditions plus favorables. En conséquence, les entreprises sont incitées à éviter de publier des pertes dans la mesure où cela leur permet de diminuer leurs « coûts de transaction » avec leurs parties prenantes (deuxième argument) qui n'ont pas la capacité, ni le temps (premier argument) d'analyser les états financiers de l'entreprise et qui vont fonder leur prise de décision sur des critères binaires comme le fait d'atteindre un seuil. À l'inverse, rater un seuil génère un coût.

La référence à la théorie des coûts de transaction dans la recherche sur les seuils appelle trois séries de remarques :

- Tout d'abord, la théorie des coûts de transaction développée par Coase (1937) et Williamson (1985) vise à expliquer la constitution de firmes intégrées par opposition au recours au marché. Autrement dit, cette théorie appartient au champ des théories de l'organisation. Dans le contexte de l'étude des seuils, c'est essentiellement la seule notion de rationalité limitée des agents de Simon (1947) qui est mise en œuvre. Les coûts de transaction évoqués par les auteurs en comptabilité dans le premier argument s'apparentent aux coûts d'agence (Jensen et Meckling, 1976). En aucun cas les auteurs ne font référence à des coûts de transaction tels qu'ils sont décrits par les auteurs en théorie des organisations[28].

[27] « Stakeholders » en anglais.

[28] S'ils utilisent un vocabulaire qui peut prêter à confusion, il faut cependant souligner que Burgstahler et Dichev (1997) ne se réfèrent pas explicitement aux auteurs des théories des organisations. Ce n'est pas le cas de Glaum, Lichtblau et Lindemann (2004).

- La seconde série de remarques concerne la nature des coûts générés par le ratage d'un seuil (second argument). Ils sont décrits comme des coûts liés à l'obtention des financements qui sont d'autant plus élevés que le résultat est faible. Ces coûts ne sont donc pas spécifiques aux seuils et s'apparentent aux motivations financières décrites par Healy et Whalen (1999). Ils ne peuvent pas non plus être assimilés aux coûts de transaction tels qu'ils sont décrits par les auteurs en théorie des organisations.

- La troisième série de remarques concerne l'articulation entre ce cadre explicatif et les facteurs appelés comportementaux. Le premier argument (le coût d'accès à l'information) explique pourquoi les investisseurs accordent une grande importance à la variable résultat, alors même que cette variable n'est pas la seule source d'information dont ils disposent sur les entreprises. Ce premier argument est donc à rapprocher de la théorie prospective et de la valeur symbolique des nombres pour expliquer pourquoi certains seuils de résultats vont jouer un rôle si important dans les prises de décision. Autrement dit, l'obtention de bonnes performances génère une économie pour l'entreprise (confiance accrue, obtention facilitée de capitaux, diminution des coûts d'agence…), mais cette économie n'est pas proportionnelle au résultat autour de certains seuils de référence à cause des décisions binaires et de l'imaginaire symbolique des agents. Les seuils sont concentrés sur le résultat comptable parce qu'un grand nombre d'actionnaires n'a pas les moyens d'accéder à d'autres sources d'informations. C'est en combinant les facteurs psychologiques (sur les investisseurs) d'une part, et la nature du coût du financement externe d'autre part que les auteurs imaginent une source de « surcoût » lié au non dépassement de certains seuils de publication.

En conclusion, l'évocation des « coûts de transaction » par les auteurs en comptabilité mérite d'être considérée avec prudence. Ces explications ne font pas, contrairement à ce que leur nom laisse penser, référence au champ des théories des organisations. Par ailleurs, elles ne distinguent pas (1) le coût d'obtention des informations par les actionnaires, qui explique la focalisation sur la variable résultat mais ne justifie pas l'existence de seuils, et (2) le coût du financement de l'entreprise, qui n'est pas spécifique aux seuils. Dans le cadre conceptuel développé dans le chapitre III, l'expression « coûts de transaction » est abandonnée pour éviter toute confusion. La mise en évidence du concept d'effet de seuil stricto sensu (voir page 71) permet de réconcilier ces deux sources de coûts.

d. L'arbitrage coût-avantage

Degeorge, Patel et Zeckhauser (1999) construisent un modèle pour expliquer que l'évitement d'un seuil correspond à un arbitrage entre un coût futur et un avantage présent. En effet, l'évitement actuel du seuil génère un transfert de résultat dans le temps (donc comptablement neutre à LT) mais génère un coût (dissimulation, risque de découverte, etc.). Le dirigeant a avantage à faire supporter le coût à l'entreprise si un bonus de rémunération lui est attribué lorsqu'il atteint le seuil.

Le modèle est intéressant car il souligne la problématique coût/avantage liée à la gestion du résultat pour éviter un seuil. Il souligne également le fait que le coût de la manipulation est (toute chose égale par ailleurs) d'autant plus élevé que la manipulation est importante. Il ne postule donc pas que le montant de la manipulation est marginal, mais permet d'expliquer que les manipulations marginales sont susceptibles de se produire plus souvent. Il n'intègre cependant pas le fait que l'évitement du seuil peut aussi générer un bénéfice pour l'entreprise : celui, par l'intermédiaire d'une information « positive » donnée au marché, d'un accès (ponctuel) moins onéreux au financement. C'est pourtant l'argument régulièrement avancé par les dirigeants pour justifier les pratiques de gestion des seuils (Graham, Harvey et Rajgopal, 2005).

e. L'influence de l'environnement

Les études ont montré que certains critères caractérisent les entreprises qui évitent les seuils. Ces caractéristiques permettent d'identifier des causes indirectes et permettent d'affiner la connaissance des causes des seuils.

Dechow, Richardson et Tuna (2003) constatent que les entreprises faiblement bénéficiaires[29] sont plutôt plus petites, plus jeunes, moins endettées, ont un niveau légèrement plus élevé d'actifs courts (circulants et trésorerie) et de dettes à court terme que les autres entreprises. Glaum, Lichtblau et Lindemann (2004) confirment l'influence des critères de taille et d'endettement sur l'amplitude des irrégularités.

Plusieurs auteurs montrent que plus la pression des marchés est forte, plus l'irrégularité est importante (Burgstahler, Hail et Leuz, 2006 ; Goncharov et Zimmermann, 2006 ;

[29] Les auteurs insistent sur le fait que les entreprises faiblement déficitaires ne sont pas représentatives des entreprises en général, et qu'il est délicat d'utiliser des indicateurs mélangeant les deux phénomènes: sur-représentation à droite, et sous-représentation à gauche. Paradoxalement, ils mettent en garde les chercheurs vis-à-vis de cet indicateur, mais l'utilisent pour mesurer l'évolution dans le temps des « pics ». Ce point est développé dans le chapitre IV « méthodologie ».

Shuto, 2008). Plus la protection des investisseurs est grande, moins les irrégularités sont prononcées (Shen et Chih, 2005), et les seuils sont moins visibles dans les pays dont les systèmes légaux sont forts (Burgstahler, Hail et Leuz, 2006). La fiscalité exerce également une influence (Coppens et Peek, 2005 ; Burgstahler, Hail et Leuz, 2006 ; Goncharov et Zimmermann, 2006).

En conclusion, il apparaît que les entreprises les plus visibles (âge, taille) ont plus de difficulté à manipuler les seuils. Par ailleurs, la forte pression exercée par les marchés confirme le rôle du surcoût d'origine financière dans le mécanisme d'évitement des seuils.

C. Les travaux étudiant les moyens et les conséquences

1. Les études des moyens mis en œuvre par les entreprises évitant les seuils

Les études sur les moyens mis en œuvre pour gérer les résultats sont nombreuses. On retiendra notamment la thèse de Mard (2002) qui traite de ce sujet dans le contexte français. Mais elles sont beaucoup moins nombreuses à s'intéresser aux moyens mis en œuvre exclusivement pour atteindre les seuils. Pourtant, dès leur article fondateur, Burgstahler et Dichev (1997) tentent de les identifier. Ils testent ainsi les deux leviers de gestion du résultat : les flux de trésorerie et les accruals.

a. Seuils et flux de trésorerie

Burgstahler et Dichev (1997) constatent un niveau de flux de trésorerie opérationnels plus élevé dans les entreprises faiblement bénéficiaires que dans les entreprises faiblement déficitaires. Le même constat est réalisé par Burgstahler et Eames (2006) auprès des entreprises qui dépassent les prévisions des analystes. Les auteurs en concluent que cela renforce la présomption de gestion du résultat pour éviter les seuils.

Cette conclusion est relativisée par Dechow, Richardson et Tuna (2003) qui rappellent qu'il existe une relation positive entre résultat et flux de trésorerie, et que les entreprises publiant un résultat bénéficiaire peuvent présenter des niveaux de flux de trésorerie supérieurs aux entreprises déficitaires sans être pour autant présumées manipulatrices.

Le modèle de Roychowdhury (2006) répond à cette critique. En effet, il ne se contente pas de mesurer le niveau « global » des flux de trésorerie, mais mesure leur niveau « anormal ». Le modèle original permet de détecter des politiques de baisse de prix temporaires qui augmentent les ventes, des augmentations de production qui diminuent les coûts des marchandises vendues, ou des réductions massives des dépenses discrétionnaires (budgets de communication, recherche et développement, distribution et administration générale). L'auteur montre ainsi que les entreprises faiblement bénéficiaires manipulent leurs flux de trésorerie opérationnels pour dépasser le seuil du résultat nul.

En conclusion, les entreprises qui sont présumées manipuler leur résultat pour éviter de publier une perte semblent le faire notamment en agissant sur les flux de trésorerie.

b. Seuils et accruals

Plusieurs auteurs ont cherché à évaluer le niveau des accruals des entreprises qui dépassent les seuils. Burgstahler et Dichev (1997) montrent que les entreprises qui ratent le seuil du résultat nul présentent un niveau d'accruals élevés. Shuto (2008) observe des résultats similaires auprès d'entreprises japonaises, mais l'auteur souligne la faiblesse de leur significativité, notamment autour du deuxième seuil (variations de résultat nulles). Dans une étude plus récente, une corrélation positive est mesurée entre plusieurs variables[30] mesurant les accruals et les entreprises qui atteignent les seuils (Ayers, Jiang et Yeung, 2006). Si les auteurs constatent une corrélation entre la performance et les accruals, cette corrélation est cependant beaucoup plus forte pour les entreprises qui se situent autour du seuil. Ce résultat conforte donc l'hypothèse d'utilisation des accruals par les entreprises qui dépassent de peu un seuil.

Les mêmes recherches ont été menées pour expliquer le troisième seuil (erreurs de prévisions). Payne et Robb (2000) montrent que les accruals discrétionnaires sont anormalement élevés le mois précédant l'annonce des résultats lorsque les résultats non manipulés sont inférieurs aux prévisions. Kasznik (1999) constate que les entreprises qui surestiment leurs résultats ont un niveau d'accruals discrétionnaires positifs significatif, et que leur amplitude est positivement corrélée aux variables d'approximation utilisées pour évaluer les coûts de litiges. Les accruals discrétionnaires calculés selon la méthode de Jones (1991) et Dechow, Sloan et Sweeney (1995) sont significativement plus élevés l'année de la prévision que les autres années, ce qui signifie que la manipulation n'est pas

[30] Dites « variables d'approximation » ou « proxies » en anglais.

liée à une caractéristique de l'entreprise mais à l'évitement du seuil. Plusieurs auteurs confirment le niveau anormal des accruals chez les entreprises qui dépassent les prévisions des analystes (Matsumoto, 2002 ; Barua, Legoria et Moffitt, 2006 ; Burgstahler et Eames, 2006). Par ailleurs, les entreprises qui arrondissent leur résultat de manière à atteindre un seuil présentent des niveaux d'accruals élevés (Das et Zhang, 2003).

Cependant, ces résultats ont été remis en cause. Dechow, Richardson et Tuna (2003) rappellent qu'il existe une relation positive entre bénéfice et accruals, donc qu'un niveau supérieur des accruals pour les entreprises faiblement bénéficiaires ne suffit pas à démontrer la gestion du résultat. Mais surtout, ils ne parviennent pas à retrouver les conclusions de Burgstahler et Dichev (1997) lorsqu'ils mesurent les seuls accruals discrétionnaires, grâce à un modèle dérivé du modèle de Jones (1991), dont le pouvoir explicatif est pourtant largement supérieur à celui de leurs précurseurs. Plus précisément, les auteurs observent que les deux groupes (faiblement bénéficiaires et faiblement déficitaires) ont des accruals discrétionnaires supérieurs à la moyenne. De plus, l'irrégularité ne semble pas décroître lorsqu'ils éliminent de leur échantillon les entreprises dont les accruals discrétionnaires sont négatifs.

Par ailleurs, contrairement aux attentes, les entreprises dont le résultat est juste inférieur aux prévisions présentent des accruals anormalement élevés (Dechow et Skinner, 2000 ; Roychowdhury, 2006). Autrement dit, l'hypothèse selon laquelle les entreprises dont le résultat (non manipulé) est faiblement négatif ou faiblement inférieur aux prévisions manipulent leurs comptes en augmentant leur niveau d'accruals discrétionnaires pour éviter le seuil n'est pas démontrée.

Dans un article critique, McNichols (2003) revisite les conclusions des tests de Dechow, Richardson et Tuna (2003) et s'interroge d'une part sur l'intérêt d'utiliser les accruals pour mesurer une gestion subtile des résultats, et d'autre part sur la pertinence d'utiliser la mesure de la discontinuité comme indicateur de gestion du résultat. Elle souligne l'intérêt de mieux comprendre le phénomène caché derrière les discontinuités de distribution des résultats, afin d'identifier dans quelle mesure il est fondé de l'interpréter comme une manifestation de la gestion du résultat ou non.

En conclusion, la piste de recherche confrontant accruals et étude des seuils fournit des résultats mitigés, et parfois contradictoires.

c. Moyens spécifiques mis en œuvre

Constatant les limites des études mesurant les accruals anormaux, McNichols (2000) suggère que l'étude d'accruals spécifiques est la voie la plus fructueuse pour étudier la gestion du résultat. De nombreux auteurs vont suivre cette recommandation. Ainsi, plusieurs auteurs montrent que les entreprises utilisent plus spécifiquement le compte impôts différés pour gérer leur résultat afin d'atteindre le seuil des prévisions des analystes (Burgstahler, Elliott et Hanlon, 2003 ; Holland et Jackson, 2004 ; Frank et Rego, 2006). Le même constat est réalisé auprès d'une population composée exclusivement de banques (Schrand et Wong, 2003). Enfin, les entreprises qui publient des petites augmentations du résultat diminuent ces années-là les montants alloués à des fondations philanthropiques (Petrovits, 2006).

Philips, Pincus et Rego (2003) vont plus loin sur cette voie. Ils étudient la corrélation entre les évitement de seuils et les trois variables (1) impôts différés, (2) accruals anormaux et (3) accruals totaux. Ils montrent que le niveau des impôts différés est l'indicateur le plus corrélé aux irrégularités. Les auteurs en déduisent donc que cet indicateur est le meilleur pour mesurer la gestion du résultat pour éviter les seuils.

d. Utilisation des arrondis

Das et Zhang (2003) montrent que la part des entreprises qui arrondissent au centime supérieur est anormalement élevée parmi les entreprises qui atteignent un seuil (résultat nul, prévision des analystes, variation nulle de résultat). Ils en déduisent que arrondir est un moyen d'atteindre un seuil. Autrement dit, les entreprises peuvent limiter le montant à manipuler de manière à permettre, grâce à l'arrondi, d'atteindre le seuil.

e. Moyens mis en œuvre pour limiter les discontinuités

Les contraintes à la gestion du résultat s'appliquent également dans le contexte des manipulations pour éviter les seuils. Elles sont de trois natures (Jeanjean, 2001) : (1) contraintes techniques (limitation des choix via la réglementation comptable, poids des décisions prises antérieurement via le principe de permanence des méthodes), (2) contrôle externe effectué par les auditeurs, et (3) contrôle exercé en interne par le gouvernement d'entreprise.

L'impact de ces contraintes a été testé empiriquement. La qualité de l'audit semble limiter la gestion du résultat pour atteindre le seuil du résultat positif, mais pas pour éviter la diminution du résultat (Caneghem, 2006). Les manipulations pour éviter de publier une

perte sont moins importantes lorsque le nombre d'heures d'audit facturées augmente (Caramanis et Lennox, 2008). Les entreprises auditées par les « Big 5 » présentent une irrégularité de distribution du résultat moins importante que les entreprises auditées par un « Non Big 5 » et les entreprises auditées par un « Big 5 spécialiste » ne présentent presque pas d'irrégularité. Par ailleurs, la présence d'un comité d'audit diminue l'amplitude des irrégularités (Davidson, Goodwin-Stewart et Kent, 2005). Enfin, Il n'y a pas, à notre connaissance, d'étude croisant contraintes techniques, et contraintes gouvernementales, avec les seuils.

2. Les travaux étudiant les conséquences des seuils

a. Les entreprises qui évitent les petites pertes

Degeorge, Patel et Zeckhauser (1999) étudient les conséquences des effets d'évitement sur les résultats futurs des entreprises. Ils mettent en évidence un résultat particulièrement intéressant : les entreprises faiblement bénéficiaires une année, donc supposées manipuler leurs comptes pour éviter une faible perte, affichent par la suite des moindres performances que les entreprises faiblement déficitaires, donc supposées non manipulatrices. Il semble donc que le fait d'afficher des faibles bénéfices ne soit, paradoxalement, pas un indicateur de bonne santé des entreprises. Ce résultat conforte l'hypothèse de la manipulation comptable puisque la manipulation (supposée à la hausse) pour dépasser le seuil une année doit être compensée les années suivantes par une baisse artificielle des résultats[31]. Le même phénomène est observé auprès des entreprises françaises cotées (Mard, 2004).

Ce résultat n'est pas confirmé par Dechow, Richardson et Tuna (2003) qui observent que les entreprises faiblement déficitaires comme les entreprises faiblement bénéficiaires présentent des résultats futurs inférieurs à la moyenne, et présentent en général une variation de résultat négative (nettement plus négative que l'ensemble). Ces auteurs comparent cependant au sein des entreprises faiblement bénéficiaires celles dont les accruals discrétionnaires sont négatifs, et celles dont les accruals discrétionnaires sont positifs. Les faibles performances futures sont surtout présentes chez les entreprises dont

[31] Degeorge, Patel et Zeckhauser (1999) utilisent l'expression « Borowing for a better today ».

les accruals discrétionnaires sont positifs (donc celles qui sont supposées gérer leurs résultats à la hausse).

Bhattacharya, Daouk et Welker (2003) mesurent un indicateur d'opacité comptable utilisant notamment l'amplitude de l'évitement des petites pertes. Ils montrent que cet indicateur est corrélé positivement au coût du financement sur les marchés, et négativement aux volumes de titres échangés. Autrement dit, les marchés sur lesquels l'opacité comptable est forte (et notamment les irrégularités de distribution) sont moins efficients.

En conclusion, l'hypothèse de gestion du résultat pour éviter le résultat nul sort renforcée de ces travaux. Les entreprises anormalement nombreuses qui publient des résultats faiblement bénéficiaires semblent privilégier un effet d'annonce à court terme qui est compensé à plus long terme par de moindres performances.

b. Les entreprises qui évitent les erreurs de prévisions

Les conséquences sur les marchés financiers des erreurs de prévisions ont très tôt attiré l'attention de la communauté financière. En 1998, Arthur Levitt (1998), le président de la commission de régulation des bourses américaines (la S.E.C.) cite dans un discours le cas d'une entreprise qui, n'atteignant pas les prévisions de un penny, perd 6 % de sa valeur boursière en un jour. Il justifie ainsi l'importance de mieux évaluer la pertinence des prévisions financières.

Si les travaux de Dreman et Berry (1995) mettent en cause la qualité des évaluations proposées par les analystes, Brown (1996 ; 1997) montre que les prévisions des analystes, même imparfaites, sont néanmoins tout à fait pertinentes. Il en conclue, à l'inverse de Dreman et Berry (1995), que les marchés ne se fient pas suffisamment aux prévisions. De telles conclusions renforcent l'intérêt à porter aux conséquences des erreurs de prévisions.

Les auteurs s'intéressent à la manière dont les marchés réagissent lorsqu'une entreprise atteint ou rate les prévisions (Defond et Park, 2001 ; Bartov, Givoly et Hayn, 2002 ; Kasznik et McNichols, 2002 ; Lopez et Rees, 2002 ; Brown et Caylor, 2005). Ils observent que les marchés sanctionnent les entreprises qui ratent les seuils, et récompensent celles qui atteignent les seuils.

Par ailleurs, les dirigeants des entreprises qui publient un résultat égal ou supérieur aux prévisions vendent davantage leurs titres que ceux des entreprises qui ratent leur objectif

(McVay, Nagar et Tang, 2006). Ce dernier résultat tend à montrer que les dirigeants sont au courant du caractère exceptionnel de la performance de leur entreprise. Ils profitent de l'effet d'annonce pour vendre des titres qu'ils estiment surcotés. Cela renforce l'hypothèse de gestion du résultat et souligne le caractère potentiellement opportuniste des dirigeants.

De même, des baisses de cours significatives sont observées lorsque les résultats sont légèrement inférieurs aux prévisions (Skinner et Sloan, 2002 et Kinney, Burgstahler et Martin, 2002). Les auteurs montrent que les conséquences des faibles erreurs sont relativement plus importantes que les conséquences des fortes erreurs. Cependant, ce résultat est à relativiser dans la mesure où les entreprises qui publient un résultat faiblement inférieur aux prévisions sont généralement celles qui ont les prévisions les plus homogènes (Kinney, Burgstahler et Martin, 2002). Autrement dit, les petites surprises surprennent tout le monde, alors que les grandes surprises ne surprennent que les plus optimistes. Il est donc difficile d'attribuer à la seule magnitude/amplitude de l'erreur la cause du surcoût. L'étude de Payne et Robb (2000) va dans le même sens.

D. Les travaux mesurant et comparant les seuils

Les premiers articles qui ont mis en évidence l'existence des seuils les ont interprété comme des manifestations de gestion du résultat. L'étude des irrégularités de distribution est dès lors apparue, au début des années 2000, comme une alternative à la mesure des accruals. Cette alternative est d'autant plus attractive qu'elle semble techniquement plus facile à mettre en œuvre comparée aux modèles complexes de mesure des accruals anormaux. De nombreuses études ont donc tenté de mesurer les irrégularités afin d'effectuer des comparaisons dans le temps -évolution- (Brown, 2001 ; Dechow, Richardson et Tuna, 2003), dans l'espace -comparaisons internationales- (Leuz, Nanda et Wysocki, 2003 ; Glaum, Lichtblau et Lindemann, 2004 ; Daske, Gebhardt et McLeay, 2006), ou entre les différents seuils -hiérarchisation- (Degeorge, Patel et Zeckhauser, 1999 ; Kasznik, 1999 ; Brown et Caylor, 2005).

Pour comparer des irrégularités, il faut être capable de les mesurer de manière fiable. Il ne suffit pas de répondre à une question binaire du type « l'irrégularité est-elle significative, oui ou non ». Les études qui font des comparaisons sont donc confrontées à un problème technique de mesure. Or la revue de littérature montre que la méthodologie de mesure des irrégularités n'est pas stabilisée.

Ce problème méthodologique étant de première importance dans le cadre de la thèse, un développement spécifique lui est consacré. Il est l'objet de la partie 2 de la thèse. Dans le cadre de la partie 1, seuls les résultats des études sont présentés.

1. Les hiérarchisations

Les études ayant tenté de hiérarchiser les seuils fournissent des résultats parfois contradictoires. Degeorge, Patel et Zeckhauser (1999) sont les premiers à observer une hiérarchie entre les seuils. Ils observent que l'évitement d'une perte est prioritaire sur les deux autres seuils, et que l'évitement d'une diminution du résultat est prioritaire sur l'évitement des erreurs de prévision.

À l'inverse, Payne et Thomas (2004) montrent que les entreprises qui publient un résultat supérieur aux prévisions ne sont pas plus valorisées si elles dépassent par la même occasion un autre seuil (résultat nul ou variation nulle de résultat). Autrement dit, dépasser les prévisions des analystes semble être pour le marché l'objectif principal. Ces résultats vont dans le même sens que ceux de Dechow, Richardson et Tuna (2003) et Brown et Caylor (2005) qui observent une évolution dans le temps des hiérarchies.

2. Les études longitudinales

Plusieurs études se sont intéressées à l'évolution dans le temps des irrégularités. La principale est celle de Brown et Caylor (2005) qui montre que la hiérarchie des seuils identifiée par Degeorge, Patel et Zeckhauser (1999) est vérifiée sur la période 1985 à 1994, mais s'inverse sur la période plus récente de 1994 à 2001. Les auteurs en déduisent que les dirigeants, influencés par l'évolution des attentes du marché des capitaux, ont réorienté leurs priorités en termes de gestion du résultat pour éviter les seuils. Dans la période charnière (1993 à 1996), l'attention des médias s'est davantage portée sur la propension des entreprises à atteindre leurs résultats que celle d'éviter les baisses de résultat.

Dans le même article, les auteurs tentent également d'évaluer la vitesse de réaction des dirigeants à la modification des priorités des attentes du marché. Ils en concluent que les dirigeants apprennent rapidement. Les marchés, depuis 1997, tendent à sanctionner de plus en plus les entreprises qui ratent le seuil des prévisions (Payne et Thomas, 2004). Ces résultats montrent que les dirigeants sont rationnels lorsqu'ils gèrent leurs résultats

pour éviter les seuils, et que les médias pourraient être la cause du déplacement des priorités des marchés de capitaux comme des dirigeants.

Dechow, Richardson et Tuna (2003) observent également l'évolution dans le temps des discontinuités. Ils constatent un déclin dans le temps de l'évitement du résultat négatif et de la faible diminution de résultat, alors que l'évitement de l'erreur de prévision des analystes augmente. Les auteurs expliquent ce phénomène par des incitations liées aux réglementations des marchés financiers. Avant 1995, pour être inscrit à la bourse de New-York[32], il était important de présenter des comptes bénéficiaires. Après 1995, cette contrainte diminue si l'entreprise présente des chiffres d'affaires, capitalisation et flux de trésorerie opérationnels dépassant certains seuils.

Cependant, comme Brown et Caylor (2005), les auteurs utilisent une mesure des irrégularités par asymétrie. McNichols (2003) souligne la fragilité de cet indicateur, facile à calculer, mais difficile à interpréter. Cette limite est développée dans la partie 2 de la thèse.

En conclusion, s'il semble que l'évolution de l'environnement économico juridique influence la présence et l'ampleur des seuils dans la publication des résultats comptables, des résultats contradictoires, et des limites méthodologiques sont soulignées. Par ailleurs, aucune étude sur l'évolution des seuils en France n'a encore été publiée.

3. Les études internationales

La comparaison entre les informations comptables publiées selon des normes différentes renvoie à la littérature foisonnante sur la comptabilité comparée. Cette littérature cherche à identifier l'influence de l'environnement culturel, juridique et politique sur les systèmes comptables.

Plusieurs études internationales ont été réalisées sur les seuils.

Leuz, Nanda et Wysocki (2003) étudient la gestion des résultats dans 31 pays (entre 1990 et 2000), et la mettent en relation avec l'environnement économique et légal des pays. Les seuils sont alors un des indicateurs de gestion du résultat. Pour les auteurs, la gestion des résultats a comme origine le conflit d'intérêts entre managers (insiders) et investisseurs (outsiders) pour le contrôle des bénéfices. Les managers tendent à cacher

[32] NYSE : New York Stock Exchange.

la performance économique réelle de l'entreprise pour s'approprier des bénéfices, pour cacher les mauvais résultats d'une période, pour constituer un fond d'opération ou pour asseoir le contrôle de la firme. Des facteurs environnementaux limitent la marge d'opération des managers : une protection légale renforcée des droits des investisseurs, un système juridique efficace[33], un marché boursier développé, et une fragmentation de la structure de la propriété. Les auteurs forment trois groupes de pays en fonction de leurs configurations institutionnelles. Groupe 1 (outsiders economies) : marché boursier développé, fragmentation de la structure de la propriété, protection juridique renforcée pour l'investisseur, haut degré de publication de l'information financière (information disclosure), système juridique efficace. Groupe 2 : petit marché boursier, concentration de la structure de propriété, faible protection légale des investisseurs, bas degré de publication d'information financière et système légal médiocre. Groupe 3 : petit marché boursier, concentration de la structure de propriété, faible protection légale des investisseurs, bas degré de publication d'information financière et système légal faible.

Les auteurs établissent un indicateur de gestion du résultat composite tenant compte de trois indicateurs de mesure des accruals et d'un indicateur d'évitement des faibles pertes. Ils constatent que la gestion de résultats est faible dans des pays du groupe 1 (pays issus de la common law). Elle est plus forte dans les pays du groupe 2 (tradition juridique romano germanique) et culmine dans les pays du groupe 3 (groupe hybride composé de pays appartenant aux deux traditions juridiques, mais dont le système juridique est déficient).

Dans le cadre de la thèse, ce qui est particulièrement intéressant, c'est l'utilisation de l'indicateur d'évitement des faibles pertes pour classer 31 pays. Cet indicateur d'asymétrie fait l'objet d'un large développement dans la partie 2 de la thèse.

L'étude de Daske, Gebhardt et McLeay (2006) confirme les résultats de Leuz, Nanda et Wysocki (2003) et montre que la gestion des résultats pour éviter les seuils est plus importante dans les pays de l'Union Européenne qu'aux États-Unis.

Brown et Higgins (2001) comparent la propension des entreprises de 12 pays (dont les États Unis) à gérer leur résultat afin d'éviter les surprises (c'est-à-dire les erreurs de prévisions des analystes). Autrement dit, les auteurs comparent l'irrégularité de distribution des écarts entre résultat publié et prévision. Ils confirment leurs trois hypothèses :

[33] Les auteurs utilisent l'expression « legal enforcement ».

H1 : Les dirigeants américains sont davantage susceptibles de publier des résultats qui surpassent légèrement les prévisions que ne le sont les dirigeants hors États Unis.

H2 : Les dirigeants américains sont moins susceptibles de publier des résultats légèrement inférieurs aux prévisions que ne le sont les dirigeants hors États Unis.

H3 : Les dirigeants américains sont davantage susceptibles de gérer les prévisions des analystes, ce qui se traduit par des estimations plus homogènes en cas de petites surprises par rapport aux grandes surprises, que ne le sont les dirigeants hors États Unis.

Glaum, Lichtblau et Lindemann (2004) comparent les seuils comptables entre entreprises américaines et allemandes. Leurs résultats ne montrent pas de différence significative entre les entreprises américaines et allemandes en ce qui concerne les seuils « résultat nul » et « variations de résultat nulles ». En revanche, les entreprises allemandes semblent moins éviter le seuil « prévisions des analystes » que les sociétés américaines. Une telle étude comparative n'a jamais été menée en France.

En définitive, toutes ces études tendent à montrer que les facteurs environnementaux ont une influence sur les pratiques de gestion du résultat pour atteindre les seuils. Cependant, ces études qui utilisent les irrégularités pour expliquer les différences entre populations sont soumises à deux importantes limites :

(1) Si les entreprises qui évitent les seuils sont supposées gérer leur résultat, il est loin d'être démontré que ces entreprises sont représentatives de l'ensemble des entreprises gérant leur résultat. Les classements fondés sur de telles mesures sont d'autant plus fragiles que par postulat, si les irrégularités sont révélatrices de manipulations, celles-ci sont généralement considérées comme étant de faible envergure.

(2) Par ailleurs, les études qui adoptent explicitement le postulat selon lequel les discontinuités sont une mesure de gestion du résultat sont celles qui utilisent des indicateurs d'asymétrie. Cette méthodologie de mesure présente d'importantes limites exposées dans le chapitre IV de la thèse. Sans chercher à anticiper sur ce point, il convient de souligner que ces indicateurs peuvent conduire à masquer l'amplitude réelle des irrégularités en ne distinguant pas l'irrégularité (généralement positive) observée à droite du seuil de celle (généralement négative) observée à gauche.

E. Synthèse de la revue de littérature et définition des axes de recherche

1. Un cadre théorique incomplet

Bien que le cadre politico-contractuel de la comptabilité n'ait jamais été évoqué directement par les auteurs pour appréhender le phénomène de manipulation pour éviter les seuils de publication, il apparaît qu'il peut contribuer efficacement à l'expliquer à condition de lui apporter plusieurs précisions. La première proposition qui évoque les motivations contractuelles des dirigeants peut être adaptée pour tenir compte du fait que les dirigeants ne sont pas motivés uniquement par leur rémunération, mais aussi par leur réputation, notamment si celle-ci est liée au dépassement d'objectifs qui constituent alors des seuils de publication. Par ailleurs, les limites imposées par les contrats de prêts de la seconde proposition constituent des seuils de publication même s'ils n'ont jamais été mis en évidence empiriquement. Enfin, les coûts politiques de la troisième proposition peuvent s'exprimer lors du franchissement de certains seuils en fonction de la nature de la réglementation, et peuvent être complétés en tenant compte de l'influence de la fiscalité.

Les auteurs en comptabilité ont, en revanche, proposé des explications spécifiques à l'évitement des seuils. Les théories comportementales fournissent une explication au comportement supposé irrationnel des investisseurs face à certains chiffres, et notamment face aux chiffres négatifs. L'existence d'un coût de financement qui augmente plus que proportionnellement lorsque le résultat manque un seuil (dénommé à tort « coût de transaction ») est évoqué. Mais ces explications se focalisent sur le comportement des investisseurs, supposés influençables par les seuils, alors que les manipulations comptables sont le fait des dirigeants. Notamment, elles ne distinguent pas l'impact psychologique sur les investisseurs de celui qui peut agir sur les dirigeants. Elles négligent également les motivations personnelles des dirigeants.

Le cadre théorique permettant de comprendre la gestion du résultat pour atteindre les seuils apparaît donc comme étant incomplet et fractionné. Un effort d'articulation des différentes explications dans le cadre plus général de la littérature sur la gestion des résultats est souhaitable. Ce travail est développé dans le chapitre III.

2. Une remise en cause du postulat de manipulation

La difficulté de confirmer l'hypothèse de manipulation pour éviter les seuils par des mesures d'accruals conduit certains chercheurs en comptabilité à s'interroger sur le fondement de cette interprétation. Les articles les plus récents semblent répondre à ce doute en montrant que les accruals spécifiques (Burgstahler, Elliott et Hanlon, 2003 ; Holland et Jackson, 2004 ; Frank et Rego, 2006 ; Petrovits, 2006) ou les activités réelles (Graham, Harvey et Rajgopal, 2006 ; Roychowdhury, 2006) justifient l'hypothèse de manipulation.

En revanche, peu de choses ont été dites ou écrites sur les montants manipulés. Le postulat de manipulation marginale n'est généralement pas remis en cause. Des recherches sur ce thème sont souhaitables.

3. Des limites méthodologiques

Plusieurs limites méthodologiques ont été soulignées pouvant remettre en cause les résultats des études sur les seuils et expliquer que certaines conclusions soient contradictoires. L'impact des choix méthodologiques sur l'identification, la mesure, et la comparaison des seuils a été jusqu'alors peu étudié. Des recherches systématiques sur ce thème sont souhaitables. Trois principales limites sont identifiées dans la littérature : le choix de la largeur des intervalles d'observation, le choix du dénominateur utilisé pour mettre le résultat à l'échelle, et les biais d'échantillonnage.

a. La largeur des intervalles d'observation

Dans leur article, Glaum, Lichtblau et Lindemann (2004) développent un argumentaire concernant le choix de la largeur (et donc du nombre) des intervalles d'observation de la distribution. L'idée est qu'il faut sélectionner un intervalle suffisamment petit pour qu'il soit confiné aux gestions du résultat liées à l'évitement des seuils, mais suffisamment grand pour que les effectifs compris dans les intervalles ne soient pas trop faibles. Les auteurs évoquent généralement une méthode mathématique pour déterminer la largeur des intervalles en fonction de la taille de l'échantillon et de la volatilité des résultats (Degeorge, Patel et Zeckhauser, 1999 ; Dichev et Skinner, 2002 ; Holland et Ramsey, 2003). La formule de calcul de référence est généralement celle de Silverman (1986) et

Scott (1992) : $2(IQR)n^{-1/3}$, où IQR est l'intervalle interquartile[34] de l'échantillon et n le nombre d'observations.

Mais, si tous les auteurs citent une règle, ils retiennent toujours, in fine, une taille d'intervalle empirique. Le problème, disent-ils, c'est que l'on ne connaît pas le nombre d'entreprises ayant géré leur résultat et que la détermination de la largeur des intervalles d'observation se fait donc empiriquement, graphiquement, de manière à maximiser l'asymétrie.

Glaum, Lichtblau et Lindemann (2004) comme Holland et Ramsey (2003) font leur étude à partir de plusieurs intervalles et comparent les résultats. Ils constatent que les résultats sont très sensibles au choix de la largeur des intervalles. Ils peuvent dans certains cas inverser les conclusions. Si la taille de l'intervalle joue autant sur les résultats, c'est qu'il y a un problème méthodologique. Le problème du choix de l'intervalle semble effectivement essentiel. Il ne peut, dans une démarche comparative, être laissé à l'appréciation subjective du chercheur.

b. Le dénominateur utilisé pour « réduire » la variable résultat

Toutes les études qui observent les discontinuités de distribution des résultats observent en réalité la distribution de la variable résultat « réduite » par un indicateur de taille permettant de mettre à l'échelle les observations. Degeorge, Patel et Zeckhauser (1999) disent ainsi « homogénéiser les entreprises », et Burgstahler et Dichev (1997) expliquent que c'est pour corriger le fait que les entreprises sont issues d'un large éventail de tailles. Le dénominateur ainsi utilisé peut être le chiffre d'affaires, le total de l'actif, ou plus couramment, la valeur de marché de l'entreprise. Cependant, l'étude de Durtschi et Easton (2005) observe trois limites liées à cette pratique.

(1) Les justifications apportées n'expliquent pas pourquoi la diversité des tailles gène l'observation empirique du phénomène. Au contraire, les auteurs démontrent que la « réduction » provoque une distorsion de la distribution sous jacente. Ils constatent en effet que la distribution de la variable résultat par action non « réduite » ne présente pas de discontinuité autour du zéro. Ils vont alors chercher à évaluer dans quelle mesure la discontinuité peut provenir, non pas d'une discontinuité du numérateur, mais d'une discontinuité du dénominateur.

[34] « Interquartile range » en anglais.

(2) Il existe un biais du survivant : la part des entreprises faiblement déficitaires pour lesquelles la valeur de l'action en début d'année (le dénominateur) est manquante est plus importante que celle des entreprises bénéficiaires. Ces dénominateurs manquants tendent donc à diminuer le nombre d'entreprises dont le résultat réduit sera situé juste en dessous du seuil. De plus, le marché n'évalue pas de la même façon les entreprises bénéficiaires (même faiblement) que déficitaires (même faiblement). Cette distorsion du dénominateur lorsque le résultat est faiblement supérieur ou faiblement inférieur au seuil peut expliquer une discontinuité de la variable résultat réduit. Les auteurs montrent que cette distorsion du numérateur est valable aussi bien pour la variable « valeur de marché », que pour la variable « total actif ».

(3) Les limites concernant le seuil « résultat nul » seraient également valables pour le seuil « variation nulle de résultat » car le signe de la variation de résultat est corrélée au niveau de résultat (Beaver, McNichols et Nelson, 2007).

Il résulte de ces observations que l'utilisation d'une variable « réduite » implique de tenir compte aussi bien du dénominateur que du numérateur lorsque l'on cherche à interpréter une discontinuité, et qu'il est important de contrôler la neutralité du dénominateur dans les études de distribution des résultats. Ce problème méthodologique a été largement sous-estimé jusqu'alors dans la littérature.

c. La composition des échantillons

Durtschi et Easton (2005) soulignent également des biais dans la composition des échantillons de la plupart des études sur les seuils. Il existe un biais du survivant concernant le dénominateur « valeur de l'action du début de l'année ». Les auteurs montrent que, en ce qui concerne les prévisions des analystes telles qu'elles sont fournies dans la base IBES, la part des entreprises suivies par les analystes faisant des pertes (même faibles) est très nettement inférieure à la part des entreprises suivies faisant des bénéfices (même faibles). L'intérêt sélectif des analystes a donc un impact sur la distribution des erreurs de prévisions, et potentiellement sur la discontinuité observée autour du seuil variation nulle de résultat.

En conclusion, que ce soit le problème de la largeur des intervalles, de l'impact du dénominateur, ou des biais de composition des échantillons, il apparaît que les études réalisées jusqu'alors sur les seuils de publication des résultats ont sous-estimé les conséquences sur les résultats des choix méthodologiques effectués. Les discontinuités

sont « trop » faciles à mettre en évidence et en quelque sorte, la « robustesse » des modèles utilisés n'est pas testée.

4. Définition des axes de recherche

En conclusion, la revue de littérature enseigne :

- Qu'il existe des irrégularités de distribution observables de manière quasi systématique, quelle que soit la population ou la période. Ces irrégularités ont été constatées en France mais ce courant de recherche est actuellement essentiellement anglo-saxon. L'article de Mard (2004) publié sur le sujet demeure une exception.

- Qu'elles sont interprétées comme la manifestation de manipulations et que cette interprétation est d'autant plus admise qu'elle est reconnue par les dirigeants. Ces manipulations sont présumées porter sur des faibles montants. Cette hypothèse n'a cependant pas été démontrée.

- Que le cadre d'analyse dépasse le champ habituel (théorie positive) de la recherche comptable sur base de données. Il est composé d'explications hétéroclites faisant appel à des théories issues de la psychologie ou de la finance comportementale.

- Que les mesures utilisées divergent pour mesurer les irrégularités ce qui limite la portée des résultats des études comparatives. La recherche sur le sujet souffre de limites méthodologiques mal étudiées.

Les deux lacunes (cadre d'analyse et méthodologie) identifiées orientent le travail de recherche et permettent de définir le sujet de thèse. L'étude sur les seuils va donc s'articuler autour de trois axes :

(1) Le premier axe consiste à préciser le cadre conceptuel, répondant ainsi à la première limite identifiée. Elle fait l'objet du chapitre qui suit (chapitre III).

(2) Le deuxième axe consiste à mieux cerner l'impact des choix méthodologiques sur les mesures des irrégularités. La partie 2 présente ces choix et teste empiriquement leurs effets sur un échantillon d'entreprises françaises cotées de 1992 à 2004.

(3) Le troisième axe vise à proposer un nouvel outil de mesure des irrégularités fondé sur une démarche paramétrique. Il est mis en application dans le cadre de l'étude de l'évitement des pertes en France par les sociétés cotées. Il conduit à réviser le postulat de manipulation marginale. Les résultats de l'étude conduisent en effet à penser que, si manipulation du résultat il y a pour éviter de publier une petite perte, cette manipulation peut porter sur des montants très importants.

CHAPITRE III - CADRE THÉORIQUE EXPLICATIF DE LA GESTION DES SEUILS

La revue de littérature a mis en évidence l'absence de cadre théorique homogène pour analyser les causes des seuils comptables. Alors que ces études semblent appartenir au courant de la recherche positive en comptabilité, aucun article n'évoque explicitement le cadre politico-contractuel de Watts et Zimmerman (1978). Seules des explications partielles ont été fournies, et leur articulation avec le cadre de la recherche sur la gestion du résultat en général n'est jamais exposée.

Ce troisième chapitre se donne comme objectif de combler cette lacune en proposant un cadre théorique général d'analyse de la gestion des seuils comptables. Pour mener à bien ce premier apport, il est indispensable de mener une réflexion sémantique préalable (section A), nécessitant une définition des concepts utilisés dans la littérature sur les seuils (discontinuité, irrégularité, seuil, effet de seuil). Une typologie des seuils comptables est proposée, et la notion d'effet de seuil est clairement isolée de la notion d'effet d'évitement. Ces précisions permettent (section B) de proposer le cadre d'analyse qui articule la littérature sur les seuils dans le cadre plus général de la littérature sur la gestion du résultat. Il croise les motivations personnelles du dirigeant et les motivations économiques pour l'entreprise, avec leur origine contractuelle, réglementaire ou financière. Il intègre le rôle des facteurs psychologiques en distinguant leur effet sur les dirigeants et sur les investisseurs. Le rôle médiateur des anticipations des dirigeants permet d'articuler l'ensemble de ces éléments.

Le chapitre III se termine (section C) par une synthèse de la partie 1 et l'énoncé définitif de la problématique de recherche développée dans les parties 2 et 3.

A. Un éclairage terminologique conduisant à distinguer deux « effets de seuils »

Un travail d'approfondissement sur la terminologie est un préalable à la compréhension des seuils comptables. Les notions de discontinuité, d'irrégularité, et de seuil sont précisées. Une typologie des seuils est proposée. Deux interprétations de la notion « d'effet de seuil » sont distinguées.

1. Discontinuité, irrégularité et seuil

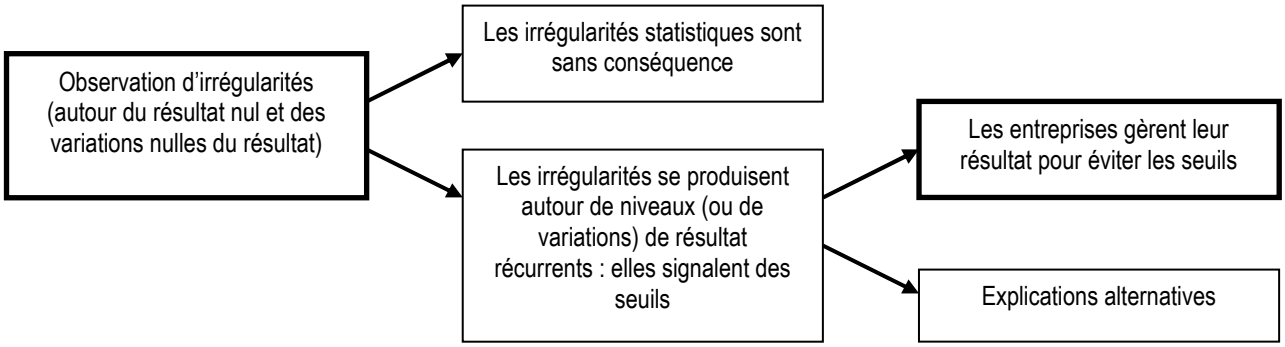
Une *discontinuité* dans une courbe de distribution des résultats n'est pas nécessairement une irrégularité. La discontinuité est un « pic », un « saut » dans la distribution. Ce n'est une irrégularité que si l'on postule que la distribution suit une loi uniforme. À l'inverse, une irrégularité n'est pas exclusivement une discontinuité. Une queue épaisse de distribution (l'effet fortes pertes, ou apurement des comptes, ou grand bain comptable[35]) est, par exemple, souvent considérée comme une irrégularité alors qu'elle n'est pas une discontinuité. Les chercheurs qui analysent les résultats des entreprises postulent que leur distribution devrait présenter une allure « plus ou moins » gaussienne en l'absence de manipulation comptable. Aucun n'affirme précisément quelle loi de distribution devrait suivre la distribution mais tous s'accordent à dire que l'allure générale de la courbe devrait être lisse, sans discontinuité. Les discontinuités observées sont donc interprétées comme étant des irrégularités.

Les *irrégularités* peuvent être interprétées de deux manières. Elles peuvent être des irrégularités statistiques sans conséquence, liées à des aléas qui n'ont pas lieu d'être interprétés comme ayant une signification particulière. Elles peuvent au contraire signaler un comportement anormal de la part des entreprises. Si les irrégularités apparaissaient de manière aléatoire sur les distributions, il ne serait pas possible de les interpréter. Or des irrégularités se manifestent systématiquement autour des mêmes niveaux de résultat. C'est pour cette raison que la seconde interprétation est privilégiée. Ces niveaux jouent

[35] Plus connu sous son expression anglaise de « Big Bath Accounting ».

un rôle spécifique. Ils sont appelés « seuils », « seuils comptables », ou « seuils de publication comptable »[36].

Figure 10 : De l'observation des discontinuités à l'évitement des seuils



Dans un souci de précision terminologique, il faut souligner que les seuils ne sont jamais « mesurés » par les chercheurs ; ils sont simplement « identifiés ». Seule la discontinuité peut être « mesurée » (le « pic » peut être plus ou moins haut ou large). Le seuil est l'endroit où la discontinuité se produit.

Les *seuils* peuvent être classés en deux catégories. Les seuils « absolus » existent indépendamment des caractéristiques ou de l'historique de l'entreprise. C'est le cas du seuil du résultat nul. Les seuils « relatifs » correspondent à des niveaux propres à chaque entreprise, et dont la valeur peut changer dans le temps. C'est le cas, par exemple, du seuil de variation nulle de résultat. Le montant à atteindre est propre à chaque entreprise et change chaque année en fonction des performances de l'année précédente.

Une seconde distinction peut être faite entre les seuils dont l'existence est toujours présente, de ceux dont l'existence même n'est pas garantie. Les seuils « résultat nul » ou « variation nulle de résultat » peuvent exister dans toutes les distributions, quelle que soit l'année ou le pays étudié. Le seuil « erreur nulle de prévision » nécessite, quant à lui, l'existence préalable de prévisions. Si ces prévisions n'existent pas, ou si elles ne sont pas connues du chercheur, le seuil ne peut être observé. Il apparaît dans ce second cas une limite méthodologique potentielle liée à des biais de sélection dans la mesure où le seuil n'est observable qu'auprès d'une sous-partie de la population.

Enfin, parmi les seuils « relatifs non systématiques », il est possible de distinguer ceux dont l'agrégation sur de larges échantillons semble aisée (dès lors que l'information est disponible) comme le fait d'atteindre ou non une prévision de résultat ou un niveau minimum de rentabilité (Yu, Du et Sun, 2006), de ceux dont l'agrégation est

[36] Les termes « targets », « benchmarks », « thresholds » sont utilisés indifféremment dans la littérature anglophone.

techniquement difficile voire impossible. C'est parce que leur observation est particulièrement délicate à effectuer que de tels seuils sont absents de la littérature empirique. Par exemple, les clauses restrictives d'endettement liées à un certain niveau de performance ou de distribution de dividendes peuvent conduire les entreprises à des manipulations comptables pour éviter de les enfreindre. De telles clauses constituent donc des seuils (1) relatifs puisque leur niveau est propre à chaque entreprise, (2) dont l'existence est incertaine puisque toutes les entreprises ne sont pas nécessairement concernées, (3) dont l'observation est difficile puisqu'une entreprise peut avoir à respecter plusieurs clauses, que leur formulation n'est pas nécessairement exprimée par rapport à un niveau de résultat, et que ces clauses ne sont pas nécessairement publiées[37]. Dans cette catégorie de « seuils relatifs non systématiques dont l'observation est difficile » peuvent être classés les objectifs personnels de performance que se fixent les dirigeants et leur implication en termes de notoriété.

Tableau 3 : Typologie des seuils et conséquences méthodologiques

	Seuil absolu	Seuil relatif
Existant pour toutes les entreprises	Seuil du résultat nul : le seuil est le même pour toutes les entreprises, et son observation statistique est aisée.	Seuil de la variation nulle de résultat : le montant du résultat à atteindre est propre à chaque entreprise, mais la variation nulle de résultat est calculable pour toutes les entreprises (sauf cas particulier des créations d'entreprise dont le résultat de référence peut cependant être le résultat nul).
Existence non systématique, mais observation directe		Seuil des prévisions des analystes ou des dirigeants : le montant diffère pour chaque entreprise, et n'existe pas nécessairement (ou n'est pas disponible pour le chercheur), mais leur agrégation ne pose pas de difficulté.
Existence non systématique, et observation statistique difficile voire impossible		Seuil correspondant à des clauses restrictives d'endettement, ou à des objectifs personnels des dirigeants : le montant diffère pour chaque entreprise, il n'existe pas nécessairement, il est difficilement accessible pour le chercheur. Ces seuils posent des problèmes méthodologiques : biais d'échantillonnage, accessibilité des données, difficulté de mesurer le degré d'implication psychologique des acteurs… Les enquêtes statistiques sur ce sujet semblent peu pertinentes.

Dresser une telle typologie présente l'intérêt de souligner les différences de traitement méthodologiques qui peuvent exister pour observer les différents seuils (voir Tableau 3). Il apparaît ainsi que les seuils relatifs, s'ils ne jouent pas nécessairement un rôle moins important dans le comportement individuel des entreprises, sont généralement plus difficiles à observer (colonne de droite du tableau). Il apparaît également que les études

[37] Ces clauses sont parfois publiées. C'est le cas notamment aux États-Unis.

de cas peuvent avoir un intérêt pour étudier les seuils relatifs non systématiques dont l'observation statistique est difficile ou impossible (case en bas à droite du tableau).

2. La notion d'effet de seuil

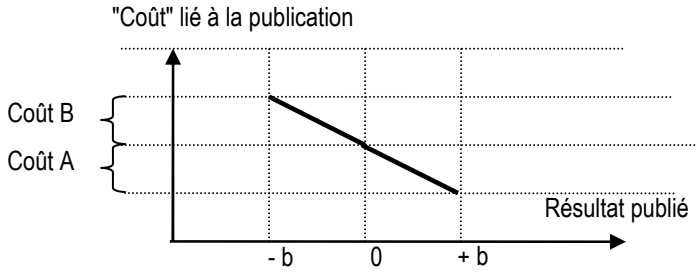
Un processus à *effet de seuil* est un processus dont les effets n'évoluent pas proportionnellement à leurs causes passé un certain seuil. Autrement dit, on parle d'effet de seuil lorsque la loi qui lie une variable (indépendante) à une autre (dépendante) va soudainement changer de nature à partir d'un certain niveau. C'est le cas de l'élastique qui se rompt passé un certain niveau de traction. En climatologie, un processus est dit à effet de seuil quand il existe un certain niveau de perturbation en dessous duquel la conséquence est proportionnelle à la perturbation, et au-delà duquel les choses se mettent à évoluer de manière totalement différente. Il est intéressant de souligner qu'un franchissement de seuil s'appelle alors une catastrophe (tornade en climatologie, tremblement de terre ou raz de marée en géologie). La notion d'effet de seuil a donc une connotation psychologique négative, et évoque le danger de ne pas pouvoir maîtriser les évènements qui risquent de se produire (Vidal, 2007).

D'autres champs utilisent la notion d'effet de seuil. En biologie, on parle d'effet de seuil lorsqu'un organisme biologique possède une réserve physiologique devant être comblée avant l'apparition d'effets délétères. Pour les substances qui entraînent ce genre de réponse, il est possible de définir une dose en dessous de laquelle les organismes exposés ne subiront pas d'effets toxiques. En informatique, les procédures étant fondées sur des décisions binaires, les effets de seuil sont fréquents, entraînant par exemple la panne générale d'un système suite à l'incrémentation marginale d'un seul paramètre. En économie enfin, il est également possible d'observer des effets de seuil. L'augmentation du taux d'imposition, par exemple, génère normalement un accroissement des ressources de l'État. Mais, passé un certain niveau de pression fiscale, l'effet d'une augmentation marginale peut avoir un effet nul, voire négatif s'il génère des effets secondaires comme le découragement ou la délocalisation des entreprises et des personnes.

Dans le cadre de la publication des résultats, parler d'effet de seuil stricto sensu signifie que passé un certain niveau de résultat publié, l'entreprise fait face à des conséquences « disproportionnées », c'est-à-dire non proportionnelles à la variation de résultat correspondante. Soit l'exemple d'une entreprise qui a le choix entre publier une petite

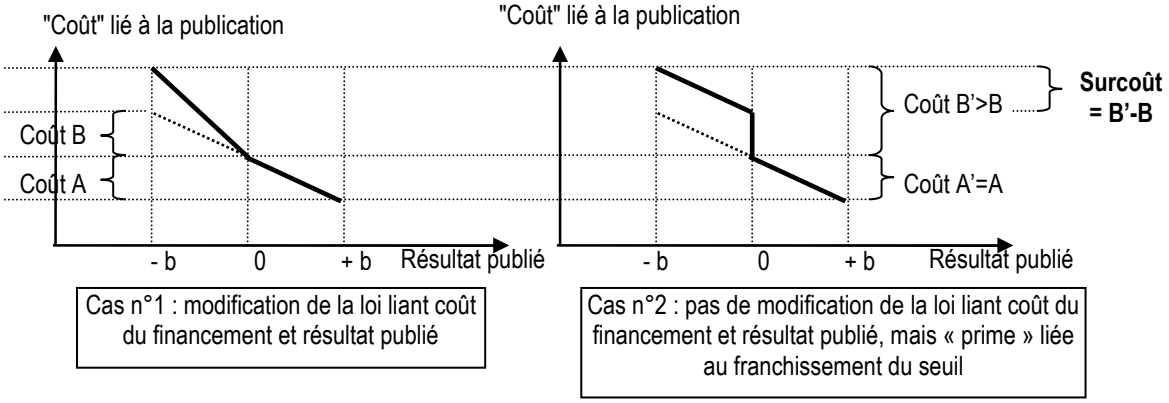
perte –b, un résultat nul, ou un petit bénéfice +b. Plus le résultat publié est faible, plus cette information jugée mauvaise est susceptible de générer un coût (la nature de ce coût est discutée dans la section B). Si l'effet de la publication du résultat est proportionnel, le coût généré par la publication d'un résultat nul au lieu de +b (coût A sur la Figure 11) doit être identique au coût généré par la publication d'un résultat négatif –b au lieu d'un résultat nul (coût B sur la Figure 11). Il n'y a pas d'effet de seuil.

Figure 11: Absence d'effet de seuil



Au contraire, si le coût généré par la publication d'un résultat -b à la place d'un résultat nul est supérieur à celui généré par la publication d'un résultat nul à la place d'un résultat +b (coût B' dans la Figure 12 supérieur au coût B de la Figure 11), il y a un effet de seuil stricto sensu. La Figure 12 présente deux situations possibles illustrant toutes deux un effet de seuil. Le surcoût peut être dû à la modification de la loi liant le coût de financement au résultat publié (cas n°1 : la pente de la courbe change) ou à une « prime » liée au franchissement du seuil (cas n°2 : la pente ne change pas, mais un « saut » est réalisé lorsque le résultat devient négatif). Dans un cas comme dans l'autre, il y a effet de seuil stricto sensu.

Figure 12 : Effet de seuil stricto sensu



Dans la littérature comptable, cette notion d'effet de seuils n'a cependant jamais été clairement définie. Au contraire, il y a une confusion dans la mesure où le terme effet de seuil peut être utilisé dans un deuxième sens : *effets provoqués par la proximité du seuil*. Le chercheur ne se situe plus ici dans la même optique objective d'identification d'effets disproportionnés passé un certain seuil, mais dans une optique subjective dans laquelle l'impact du seuil sur les acteurs, et notamment sur les dirigeants se manifeste avant même que le seuil ne soit franchi. Ainsi, si l'effet de seuil stricto sensu décrit un surcoût généré par le non-évitement d'un seuil, l'effet de seuil au sens large décrit le comportement d'évitement généré par la proximité d'un seuil (c'est-à-dire la manipulation). La terminologie est donc pernicieuse car sous la même expression peuvent se confondre deux effets distincts, qui peuvent se combiner, ou exister indépendamment l'un de l'autre (Tableau 4). C'est pourtant la distinction entre les deux interprétations du terme générique d'effet de seuil qui permet d'incorporer au cadre théorique les éléments de psychologie via la prise en compte des anticipations des dirigeants (aspect développé dans la section B).

Tableau 4 : Liens entre effet de seuil stricto sensu et effet d'évitement

	Existence d'effet de seuil stricto sensu (le coût n'est pas proportionnel au résultat)	**Absence d'effet de seuil stricto sensu** (le coût est proportionnel au résultat)	
Existence d'effet de seuil au sens large, ou **effet d'évitement** (la proximité d'un seuil incite à la manipulation)	L'effet de seuil qui génère un surcoût peut justifier l'effet d'évitement	Le surcoût ne justifie pas l'effet d'évitement. Il faut chercher d'autres motivations	Existence d'une discontinuité statistique
Absence d'effet d'évitement (= pas de manipulation)	L'entreprise n'évite pas le surcoût		Absence de discontinuité statistique

3. De l'observation des discontinuités à la gestion du résultat

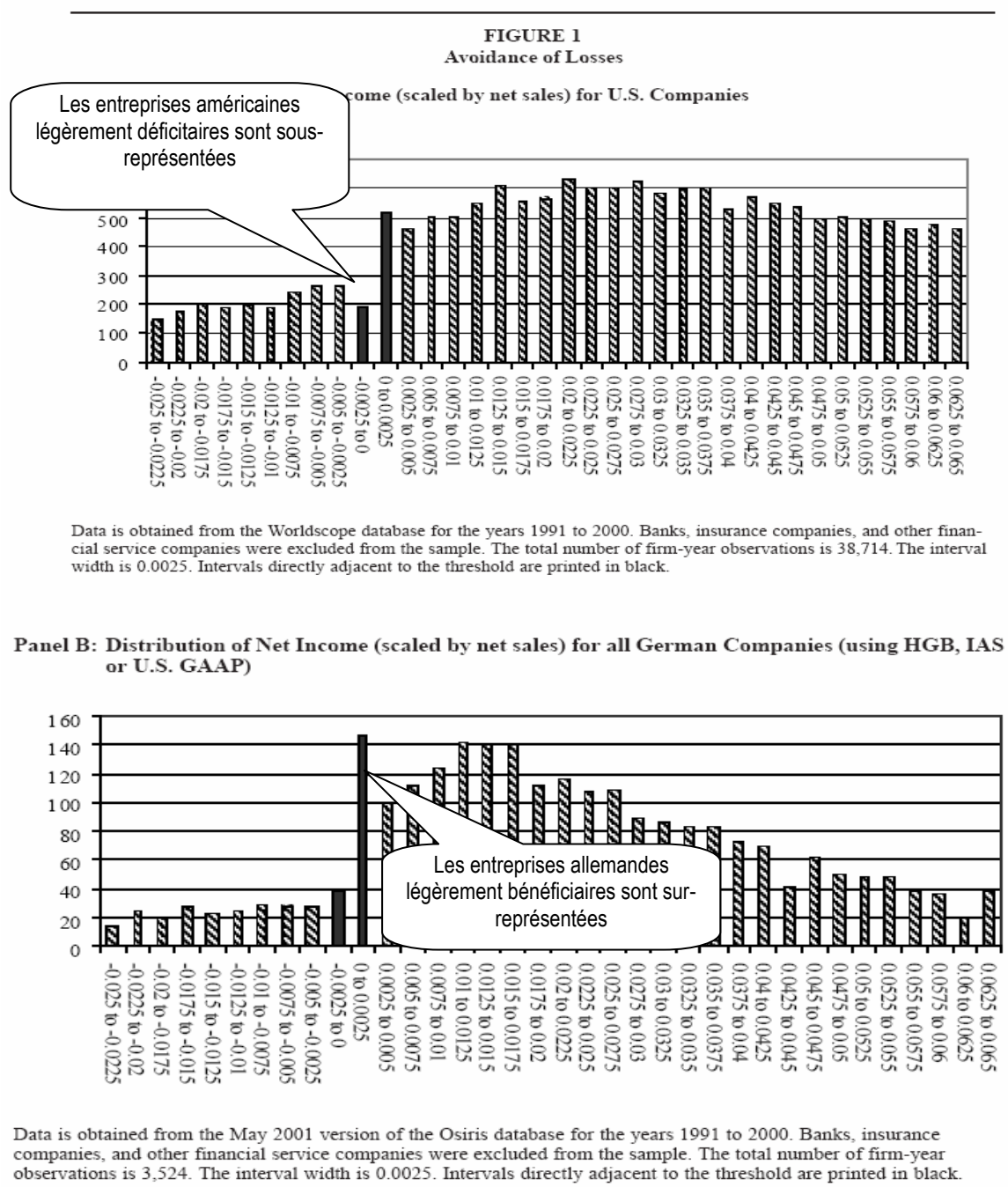
Dans un souci de précision terminologique, il est utile de rappeler l'enchaînement logique qui conduit les chercheurs à interpréter les seuils comme des manifestations de gestion du résultat. Il est notamment intéressant de souligner un constat rarement exploité dans les études sur les seuils. Chaque discontinuité observée présente en réalité une double irrégularité (voir Figure 7) : une sur-représentation des effectifs situés légèrement au-dessus du seuil, et une sous-représentation des effectifs situés légèrement en dessous du seuil.

L'explication proposée par Burgstahler et Dichev (1997) de ce phénomène de double irrégularité repose sur l'hypothèse que les entreprises dont le résultat (avant manipulation) est inférieur au seuil gèrent ce résultat pour dépasser le seuil. La manipulation du résultat serait d'autant plus tentante qu'elle est alors difficilement détectable. Ce postulat de manipulation marginale est repris par l'ensemble des chercheurs en comptabilité à l'exception de Degeorge, Patel et Zeckhauser (1999). Cependant, dans l'état actuel de la recherche sur le sujet, force est de constater que la provenance des entreprises surnuméraires qui publient des résultats légèrement supérieurs à un seuil est inconnue, comme la destination des entreprises manquantes en dessous des seuils.

L'intérêt de souligner l'existence d'une double irrégularité autour des seuils permet d'attirer l'attention des chercheurs sur la possibilité d'étudier séparément chacune de ces irrégularités. Ainsi, les études qui mesurent l'irrégularité en utilisant un ratio d'asymétrie[38] présupposent que les deux irrégularités sont le résultat du même phénomène (Leuz, Nanda et Wysocki, 2003 ; Glaum, Lichtblau et Lindemann, 2004 ; Brown et Caylor, 2005 ; Burgstahler, Hail et Leuz, 2006). Pourtant, l'observation graphique des résultats de Glaum, Lichtblau et Lindemann (2004) laisse penser que la non significativité des tests de comparaison des entreprises allemandes et américaines masque une double divergence. Les entreprises allemandes légèrement bénéficiaires sont visiblement surreprésentées, alors que les entreprises américaines légèrement déficitaires sont visiblement sous-représentées. L'indicateur composite annule, par compensation, les deux différences. La mise en lumière de cette double irrégularité ouvre de nouvelles perspectives de recherche.

[38] Que ce soit l/r ou (l-r)/(l+r), avec « r » l'effectif de la classe juste supérieure au seuil, et « l » l'effectif juste inférieur.

Figure 13 : Comparaison des distributions des entreprises cotées américaines et allemandes (Glaum, Lichtblau et Lindemann, 2004) : mise en évidence de deux irrégularités distinctes



Pour retenir l'interprétation de manipulation marginale, il faudrait être capable de comparer le nombre d'entreprises surnuméraires à droite, et le comparer au nombre d'entreprises manquantes à gauche. Si ces deux nombres sont égaux, alors l'hypothèse selon laquelle les entreprises juste à gauche manipulent pour se retrouver juste à droite est confirmée. Mais si les deux nombres ne coïncident pas, ce qui graphiquement semble être généralement le cas, alors l'hypothèse de manipulation marginale doit être remise en cause.

B. Proposition d'un cadre d'analyse

Dans le cadre explicatif des manipulations comptables spécifiquement liées à l'évitement des seuils de publication proposé dans cette section, les anticipations du dirigeant jouent un rôle médiateur. Pour Breton et Stolowy (2004), la manipulation des comptes est l'activité du dirigeant. Les autres acteurs n'agissent pas directement sur les comptes. Ils tentent éventuellement d'influencer le dirigeant pour qu'il manipule ou ne manipule pas. Les auteurs en déduisent qu'une *théorie des manipulations comptables* ne doit pas se référer aux techniques mises en place, mais aux besoins, opportunités, et relations entre les différentes catégories de parties prenantes. C'est ce point de vue qui est adopté. Les dirigeants sont les responsables de la publication des comptes et la manipulation pour éviter un seuil se fait à leur initiative ou sous leur contrôle. Il n'y aura donc manipulation que si les dirigeants voient un intérêt à le faire.

Le dirigeant peut chercher à agir dans son intérêt propre (motivations opportunistes), mais il peut chercher à agir dans l'intérêt de l'entreprise elle-même, ou de certaines des parties prenantes (motivations économiques). Dans ces derniers cas, c'est l'idée qu'il se fait de l'intérêt des parties prenantes qui le pousse à agir. Comme les facteurs psychologiques jouent un rôle important, souligner que les pratiques de gestion du résultat pour éviter les seuils sont liées à la formation du jugement du dirigeant (ses anticipations) permet d'articuler l'ensemble des éléments théoriques abordés jusqu'alors dans la littérature sur les seuils.

1. Les motivations économiques de l'évitement des seuils

La première source de motivation peut découler de la connaissance par le dirigeant de l'existence d'un surcoût généré par le non-évitement d'un seuil. Cette motivation se fait dans l'intérêt économique de l'entreprise et des actionnaires (dont le dirigeant fait peut-être partie). Le cadre de Healy et Whalen (1999) peut être utilisé pour classer les origines de ce surcoût :

(1) *Le marché* : Dans un contexte de financement par le marché, le résultat est une information destinée en priorité aux investisseurs, et plus spécifiquement aux actionnaires. Le résultat mesure l'accroissement du patrimoine des actionnaires, et la

capacité de l'entreprise à distribuer des dividendes. Dans le cadre de la théorie financière, la publication d'un résultat, a des conséquences directes sur les cours boursiers et la fortune des propriétaires (et des dirigeants si leur rémunération dépend des performances boursières). Une entreprise qui publie un résultat décevant sera boudée par les actionnaires qui demanderont une prime de risque pour investir. Le coût d'accès aux financements augmente lorsque le résultat diminue. Mais ce lien entre résultat et coût du financement n'est pas linéaire aux alentours du résultat nul (Brown et Caylor, 2005) : les marchés sanctionnent les entreprises qui ratent les seuils, et récompensent celles qui atteignent les seuils. De plus, les entreprises profitables (avant manipulation) subissent des coûts plus élevés que les entreprises déficitaires lorsqu'elles ratent le seuil des prévisions des analystes (Hayn, 1995 ; Barua, Legoria et Moffitt, 2006).

(2) *La réglementation* : Dans certains secteurs (notamment dans le secteur financier), une réglementation spécifique peut être contraignante vis-à-vis des entreprises qui franchissent certains seuils (par exemple un résultat nul ou un niveau minimum de rentabilité). La réglementation fiscale peut également jouer un rôle important lorsque le résultat fiscal est lié au résultat comptable, dans la mesure par exemple où franchir le seuil du résultat nul génère un impôt en l'absence de report de déficit. De telles contraintes juridiques incitent les entreprises à éviter ces seuils pour éviter des surcoûts d'origine réglementaire.

(3) *Les contrats* : Certaines clauses contractuelles, notamment dans le cadre des contrats d'endettement, peuvent conditionner le coût du financement à la performance de l'entreprise, et notamment à la réalisation de certains niveaux de bénéfice. Un résultat même faiblement inférieur à de tels niveaux peut générer des surcoûts d'origines contractuelles.

Par ailleurs, manipuler le résultat génère des coûts. La manipulation doit être cachée des auditeurs ou des analystes et nécessite la mise en place de montages plus ou moins perfectionnés. Elle génère par ailleurs un coût futur lié au transfert temporel des charges et produits manipulés. Elle génère enfin un coût potentiel lié au risque d'être découvert, risque de procédures judiciaires de la part des actionnaires ou de la part des autorités gérant les marchés. Éviter un seuil nécessite une manipulation du résultat de faible ampleur si le résultat non manipulé est situé faiblement en dessous du seuil. Plus l'écart entre le résultat non manipulé et le seuil est grand, plus les coûts liés à la manipulation augmentent. Il est présumé que l'augmentation de ces coûts n'est pas proportionnelle à cet écart, mais qu'elle augmente d'autant plus vite que l'écart est important (le risque augmente non linéairement). Remarquons enfin que ce coût sera d'autant plus faible que

l'entreprise, par exemple suite à une période faste, aura constitué préalablement des « réserves » en masquant une partie des bénéfices passés (gestion du résultat à la baisse via notamment la constitution généreuse de provisions etc.).

En définitive, éviter un seuil permet d'éviter un surcoût économique pour l'entreprise tout en générant un coût d'évitement. Dans une logique « coût-bénéfice », trois propositions sont formulées pour expliquer les pratiques de manipulation comptable pour éviter un seuil. P1,2,3 : Les dirigeants manipulent les résultats pour atteindre ou dépasser un objectif (seuil) si le coût généré par la manipulation demeure inférieur au gain attendu par l'entreprise, c'est-à-dire au surcoût d'origine financière (P1), contractuelle (P2) ou réglementaire (P3) généré par la publication d'un résultat inférieur au seuil.

2. Les motivations opportunistes

Parallèlement aux motivations économiques en faveur de l'entreprise, le dirigeant peut également avoir d'autres sources de motivation, plus personnelles (opportunistes) :

- *La rémunération* : L'intérêt personnel du dirigeant peut avoir une origine contractuelle, plus ou moins formelle. De manière formelle, le dirigeant peut voir sa rémunération conditionnée à la réalisation de certains objectifs, et un résultat positif ou une croissance continue du résultat peuvent être un de ces objectifs. De manière moins formelle, un dirigeant peut s'être engagé à redresser ou améliorer une situation financière (variation positive de résultat), ou à verser des dividendes (résultat positif), objectifs qui ne sont pas atteints si ces seuils ne sont pas dépassés.

-*La notoriété* : La motivation du dirigeant peut être plus informelle encore. En termes de notoriété personnelle, et de gestion de carrière, le dirigeant a intérêt à donner une bonne image de sa gestion. Dépasser régulièrement les seuils peut être un élément de construction d'une notoriété personnelle de la part du dirigeant. Une telle motivation peut être, dans certains contextes, beaucoup plus importante que des motivations contractuelles de rémunération.

Deux nouvelles propositions sont formulées. P4,5 : Les dirigeants manipulent les résultats pour dépasser un seuil si leur réputation (P4) ou leur rémunération (P5) est fortement liée au dépassement d'un objectif de résultat. Notons que ces propositions peuvent être exprimées de manière « symétriques » comme le fait Kasznik (1999) en mettant l'accent, non pas sur l'intérêt à manipuler du dirigeant, mais sur le risque de ne pas manipuler. Par exemple, plus le risque d'action en justice de la part des investisseurs est élevé en cas de

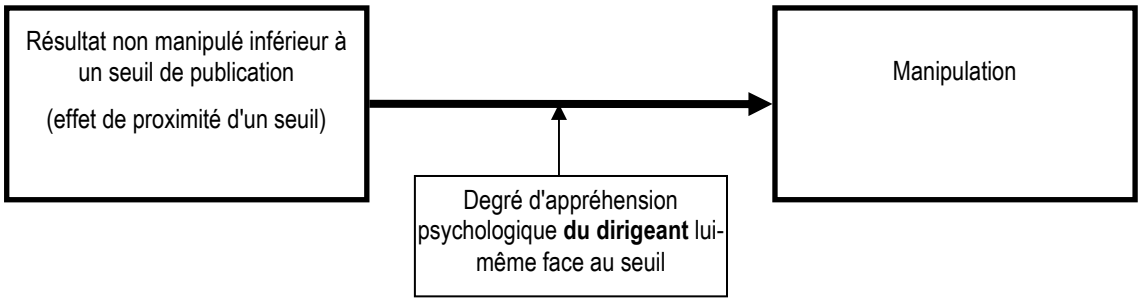
non dépassement d'un seuil comptable, plus le dirigeant choisit les méthodes comptables qui augmentent le résultat pour éviter ce seuil.

3. Le double impact des facteurs psychologiques

Comme ce sont les dirigeants qui établissent les comptes qui seront publiés, il faut distinguer l'impact des facteurs psychologiques sur les dirigeants eux-mêmes (Figure 14), de celui sur l'ensemble des investisseurs (Figure 15).

Il est légitime de se demander si les dirigeants sont eux-mêmes influencés par des facteurs psychologiques d'évitement des seuils. Mieux informés de la santé réelle de leur entreprise, et sans doute mieux formés à l'interprétation des données comptables, les dirigeants devraient accorder moins d'importance à la seule variable résultat que l'ensemble des investisseurs. Mais le degré de familiarité des dirigeants avec les chiffres comptables dépend de leur formation ou de leur environnement. Des dirigeants de PME autodidactes en matière de gestion pourraient par exemple être plus influencés par les seuils que des dirigeants de grandes entreprises ayant suivi des formations supérieures et mieux conseillés. Autrement dit, l'impact des facteurs psychologiques sur les dirigeants, s'il existe, doit être considéré comme une variable modératrice. Formulation sous forme de proposition (P6) : plus le dirigeant est susceptible d'avoir accès à des informations sur la santé de son entreprise, et/ou plus il est rodé à l'interprétation des comptes, moins il est susceptible de manipuler les comptes pour éviter les seuils.

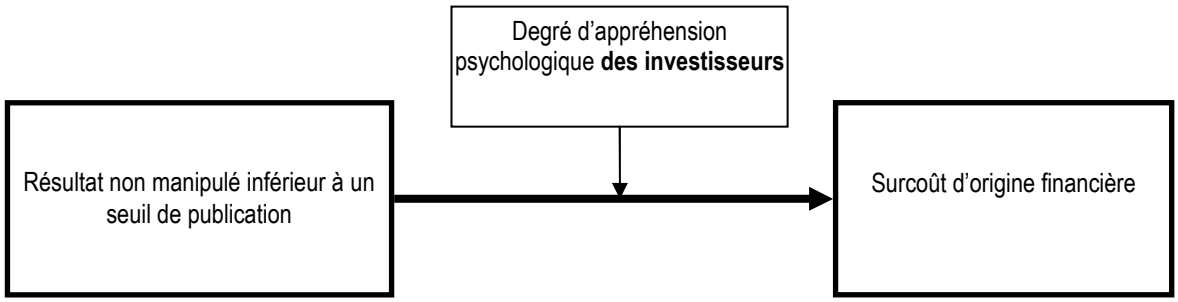
Figure 14 : Les facteurs psychologiques comme variables modératrices : effet sur le dirigeant



Parallèlement à l'effet psychologique sur les dirigeants, il existe un effet psychologique sur les investisseurs. Il est important de souligner que cet effet n'agit que sur le surcoût d'origine financière (effet de seuil stricto sensu). Autrement dit (P7), plus les investisseurs appréhendent les chiffres négatifs, plus le surcoût financier sera élevé. Soulignons enfin que l'impact peut être différent selon les catégories d'utilisateurs étudiés. Par exemple, il

apparaît que les analystes financiers anticipent la gestion du résultat pour éviter les résultats nuls, mais pas les variations nulles de résultat (Burgstahler et Eames, 2003).

Figure 15 : Les facteurs psychologiques comme variables modératrices : effet sur les investisseurs



En conclusion, les facteurs psychologiques agissent comme des variables modératrices. Les effets psychologiques sur les dirigeants n'ont jamais été mesurés empiriquement, alors que de nombreux travaux en finance comportementale étudient l'impact des facteurs psychologiques sur les investisseurs (Vidal, 2007). Mais il est important de souligner que ces derniers n'ont pas d'effet direct sur l'évitement des seuils s'ils ne sont pas anticipés par le dirigeant.

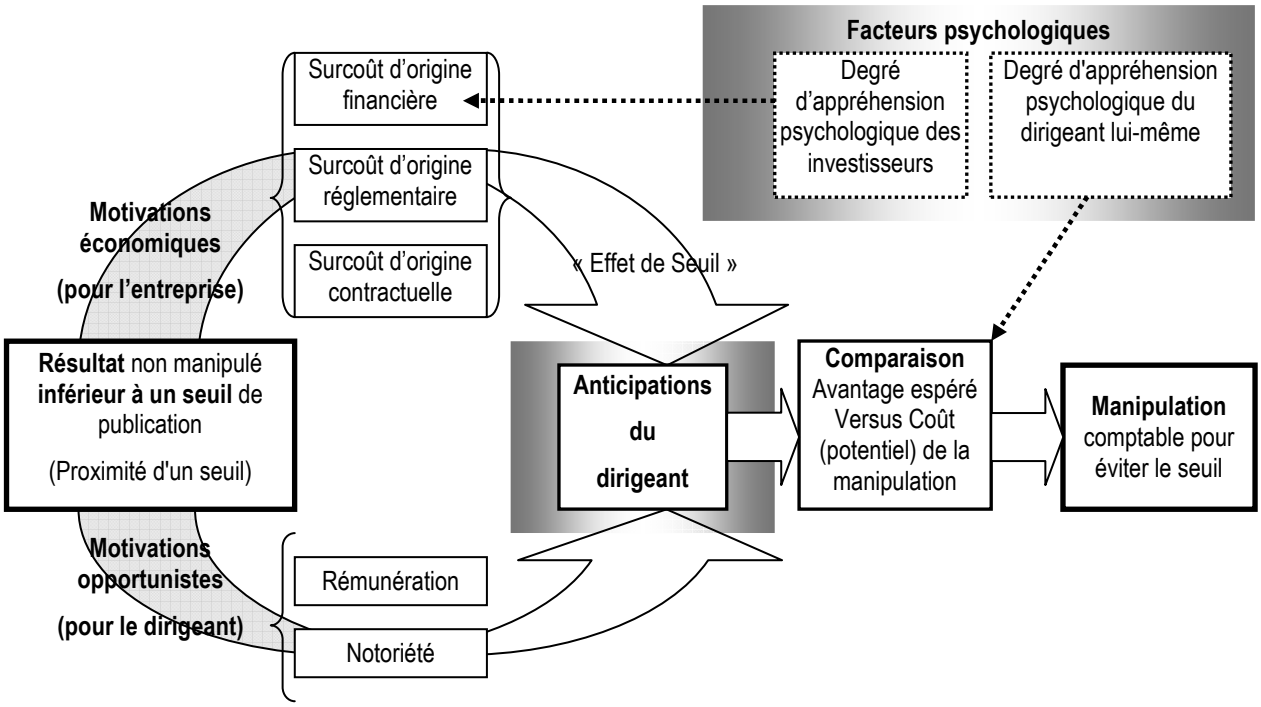
4. Le rôle des anticipations

Les facteurs psychologiques agissant sur les investisseurs eux-mêmes ont un impact direct sur les décisions individuelles de ces investisseurs, mais n'ont pas d'impact direct sur les manipulations comptables. C'est parce que les dirigeants anticipent que les utilisateurs des données comptables vont se focaliser sur l'atteinte et le dépassement de certains seuils, et qu'il ne faut pas décevoir les parties prenantes sous peine de voir certains coûts augmenter, qu'ils peuvent être incités à manipuler le résultat. Mais les anticipations des dirigeants peuvent être infondées. Les dirigeants peuvent anticiper une focalisation du marché sur le dépassement de certains seuils alors que cette focalisation, ou n'a pas lieu dans la réalité, ou n'est pas suffisante pour générer un surcoût sensible. Degeorge, Patel et Zeckhauser (1999) précisent cependant que l'effet de seuil pour l'entreprise peut être important même si peu d'acteurs sont sensibles au dépassement des seuils. Il n'en demeure pas moins que, ce qui importe, c'est moins l'existence d'un effet de seuil, que l'opinion du dirigeant sur l'existence ou non de cet effet de seuil. Dans le modèle, si les facteurs psychologiques sont considérés comme une variable

modératrice, les anticipations du dirigeant sont considérées comme une variable médiatrice.

Le rôle clef des anticipations du dirigeant permet de faire le lien entre les facteurs comportementaux et les facteurs économiques. Il a plusieurs conséquences importantes. Tout d'abord, il permet de mieux comprendre les phénomènes de hiérarchie des seuils, et la modification de cette hiérarchie dans le temps. Notamment, lorsque les médias focalisent leur attention sur certains seuils, les dirigeants constatent et suivent cette évolution en anticipant une attitude des marchés en faveur de ces seuils, modifiant ainsi les hiérarchies. Il permet ensuite de poser l'hypothèse que, quand bien même la manifestation de seuils se fonde initialement sur une cause psychologique naturelle, son expression peut évoluer par la suite en fonction du degré d'information des acteurs. La nature de l'actionnariat de l'entreprise, ou le degré d'éducation comptable des acteurs, peut influencer l'amplitude des seuils. Il permet enfin de penser qu'une meilleure information sur les seuils, la démonstration de leur existence et de leurs effets, peut conduire à leur atténuation. Autrement dit, le développement de la littérature sur les manipulations de résultat pour atteindre les seuils participe à la disparition de ces manipulations.

Figure 16 : Les motivations à la gestion du résultat pour éviter les seuils : un cadre d'analyse



5. Synthèse et apports du cadre théorique

Le cadre théorique d'analyse des seuils de publication comptable permet de mieux appréhender certains résultats secondaires figurant dans des études publiées, mais non interprétés. Par exemple, les critères de taille, de secteur d'activité et d'endettement semblent jouer sur l'amplitude des irrégularités (Dechow, Richardson et Tuna, 2003 ; Bisson, Dumontier et Janin, 2004 ; Glaum, Lichtblau et Lindemann, 2004). Une relecture de ces résultats peut être faite à l'éclairage du cadre d'analyse proposé. Il est notamment intéressant de comparer les surcoûts d'origine financière, réglementaire ou contractuelle selon la taille, le secteur ou l'endettement des entreprises.

Par ailleurs, la typologie conduit à prédire l'existence de certaines pratiques de gestion du résultat pour éviter des seuils non systématiques (notamment d'origine contractuelle ou réglementaire) encore jamais mis à jour empiriquement.

De même, l'influence des facteurs psychologiques sur les dirigeants, et le rôle des anticipations concernant les effets de seuil dans la formation de leur jugement méritent d'être étudiés. Ces facteurs suffisent-il à expliquer les divergences en matière de gestion du résultat pour éviter les seuils ? Le cadre théorique proposé permet d'envisager de nouvelles perspectives d'études sur ces sujets.

Enfin, le dernier apport n'est pas nécessairement le moindre. Si les effets d'évitement sont des irrégularités liées à des appréhensions naturelles, de nature symbolique et/ou psychologique, les identifier, les analyser, et communiquer sur leurs causes peut permettre de démystifier ces seuils, aussi bien auprès des dirigeants que des investisseurs, et par voie de conséquence, de participer à leur atténuation[39].

[39] En finance, une irrégularité (phénomène observé qui déroge à une loi) peut soit être interprétée comme un phénomène persistant qui oblige les chercheurs à modifier leurs modèles théoriques, soit comme un phénomène irrationnel dont l'observation et l'étude va entraîner la disparition.

C. Conclusion de la partie 1 : problématique de la thèse

La recherche menée dans le cadre de la thèse s'intègre dans le cadre général de la qualité des informations comptables, cadre qui résonne avec l'actualité depuis les scandales financiers de ces dernières années d'une part, et l'accélération du processus d'harmonisation internationale des normes comptables d'autre part. Ainsi, si les irrégularités de distribution sont des indicateurs de mauvaise qualité de l'information comptable, leur étude doit permettre d'améliorer la connaissance sur le thème plus général de la qualité de l'information comptable.

La recherche en gestion prend parfois des tournures d'enquête telle qu'on peut les lire dans les romans policiers. C'est le cas du courant de recherche sur les seuils de publication comptable, qui est un champ d'investigation riche et en plein développement. L'enquêteur doit identifier la *victime*, *l'arme du crime*, le *coupable*, et son *mobile* pour résoudre le mystère. La revue de littérature menée dans le chapitre II permet de préciser certains éléments de l'enquête.

La qualité du résultat est *la victime* présumée, entraînant avec elle tous les utilisateurs potentiels de l'information comptable. L'idée que le résultat comptable puisse être manipulé est d'autant plus inquiétante qu'il joue un rôle informationnel de plus en plus important dans nos économies fondées sur l'ajustement par les marchés, et le moindre indice d'irrégularité est immédiatement l'objet de toutes les suspicions. Dans ce contexte, les irrégularités de distribution statistiques, observables de manière quasi systématique, quelle que soit la population ou la période, ont été immédiatement interprétées comme une défaillance de l'outil comptable. Pourtant, aucun fondement théorique ne permet de connaître l'allure de la loi de distribution, et par ailleurs, la mesure des discontinuités est sujette à des limites méthodologiques importantes.

Les coupables présumés sont les entreprises surnuméraires situées au dessus des seuils.

L'arme du crime, c'est la manipulation comptable qui permet une amélioration du résultat, arme d'autant plus pernicieuse qu'elle serait de faible envergure donc difficile à identifier. Cette hypothèse n'a cependant pas été démontrée.

Enfin, *le mobile du crime* repose sur la présomption, sinon de malhonnêteté, du moins d'opportunisme des entreprises et/ou de leurs dirigeants qui pourraient avoir intérêt à éviter certains seuils de publication.

La réflexion conceptuelle menée dans la première partie de la thèse analyse et décortique le quatrième aspect de l'enquête : le mobile. Le cadre de la théorie politico-contractuelle (théorie positive), habituellement mobilisé dans les recherches comptables sur base de données, ne permet pas d'expliquer à lui seul l'existence d'effets de seuil. Il doit être complété par des théories issues de la psychologie, et par l'étude de la formation du jugement des dirigeants. En proposant un cadre explicatif des motivations qui peuvent pousser les entreprises à manipuler leurs résultats pour éviter des seuils de publication, la première partie de la thèse fournit aux chercheurs un cadre théorique à l'intérieur duquel il devient possible de se focaliser sur les problèmes méthodologiques de mesure des irrégularités desquels dépendent l'identification de la victime, du coupable, et de l'arme du crime. C'est ce que développent les parties 2 et 3 de la thèse.

Contrairement à la partie 1 (entièrement théorique), les parties 2 et 3 sont théoriques et empiriques. L'étude des entreprises cotées françaises permet d'illustrer les propositions développées.Plus précisément, la partie 2 étudie les limites méthodologiques et la partie 3 propose un nouvel outil d'analyse permettant de répondre à ces limites. La mise en application de cet outil dans le cadre d'une étude longitudinale de l'évitement des pertes en France compléte ainsi les travaux antérieurs essentiellement anglo-saxons.

PARTIE 2 : IMPACT DES CHOIX MÉTHODOLOGIQUES

La revue de littérature (développée dans la partie 1) a mis en évidence plusieurs limites méthodologiques. Ces limites ont peu d'impact lorsqu'il s'agit de mettre en évidence les irrégularités qui sont manifestes. Mais elles ne peuvent être ignorées lorsqu'il s'agit de mesurer et comparer les irrégularités, dans le temps ou dans l'espace.

La partie 2 de la thèse se propose d'examiner (A) l'impact du choix de la méthode retenue pour mesurer les irrégularités, (B) l'impact du choix de la largeur de l'intervalle d'étude, et (C) l'impact du choix du dénominateur. Le chapitre IV présente les enjeux théoriques de chacun de ces choix en s'appuyant sur une revue de littérature spécifique. Le chapitre VI met en oeuvre une application empirique afin de mettre en évidence l'impact réel de ces limites. Entre ces deux chapitres, le chapitre V présente les caractéristiques de la population d'entreprises cotées françaises sur laquelle les études empiriques sont menées.

CHAPITRE IV – ÉTAT DES LIEUX ET PRÉSENTATION DE L'ÉTUDE MÉTHODOLOGIQUE

La partie 2 de la thèse cherche à évaluer l'impact des choix méthodologiques sur l'identification, la mesure, et la comparaison des seuils comptables de publication. Le travail s'articule autour de trois grandes questions qui structurent la deuxième partie de la thèse :

1. Pour identifier les discontinuités, les chercheurs mesurent l'irrégularité et testent sa significativité. Pour effectuer cette mesure, plusieurs méthodes ont été utilisées. La première question de recherche porte sur l'impact de ce choix méthodologique : Quel est l'impact de la méthode utilisée pour mesurer les irrégularités ? L'objectif est de classer les mesures (non paramétriques) recensées dans la littérature en fonction de leur pertinence.

2. Mais quelle que soit la méthode retenue, il semble que les mesures sont fortement dépendantes de la largeur des intervalles d'observation. Il en découle une seconde question : Quel est l'impact du choix de la largeur des intervalles d'observation ? L'objectif est de proposer un nouvel indicateur mesurant une dimension jusqu'alors ignorée par les chercheurs : la largeur (ou « persistance ») de la discontinuité. Cette mesure permet d'estimer les montants manipulés.

3. Plusieurs variables ont servi, dans la littérature, à mettre à l'échelle le résultat étudié. Or seules les discontinuités de la variable résultat ont été interprétées, ignorant l'impact d'éventuelles discontinuités du numérateur. Il en découle une troisième question : Quel est l'impact du choix de la variable choisie pour mettre à l'échelle les résultats ? L'objectif est d'évaluer si l'impact de la variable de mise à l'échelle suffit à remettre en cause les résultats des études menées jusqu'alors.

Ces trois grandes questions sont développées dans les sections A, B et C du chapitre IV, puis font l'objet d'une étude empirique dans le chapitre VI. Par souci de clarté, Le plan est construit symétriquement. Ainsi, la section A du chapitre VI. répond à la question posée dans la section A du chapitre IV, etc.

A. Impact de la méthode de mesure de l'irrégularité

1. Présentation du problème

L'observation des irrégularités est assez simple à réaliser. En revanche, comparer des irrégularités est chose moins aisée. En effet, il est différent de répondre aux deux questions suivantes : (1) « y a-t-il irrégularité ? » et (2) « une irrégularité observée est elle plus ou moins importante qu'une autre ? ». Dans le premier cas, il suffit de constater une irrégularité, ou graphiquement, ou par un test statistique. Dans le second cas, il ne suffit plus de répondre de manière binaire « l'irrégularité est suffisamment importante, oui ou non ». La comparaison nécessite un instrument de mesure de l'irrégularité qui soit suffisamment fiable pour être comparable. Elle doit nécessairement s'affranchir d'approximations trop grossières

Dans la littérature, la méthodologie de mesure des irrégularités apparaît relativement simple. Pourtant d'importantes divergences existent entre les auteurs. Le premier axe de recherche va donc comparer les différentes méthodes de mesure utilisées par les chercheurs. Il est possible de les classer en deux grandes familles :

- Des mesures nécessitant de connaître un effectif attendu (théorique), pour estimer l'irrégularité par différence avec l'effectif observé. Ce sont des mesures d'irrégularité à proprement parler.

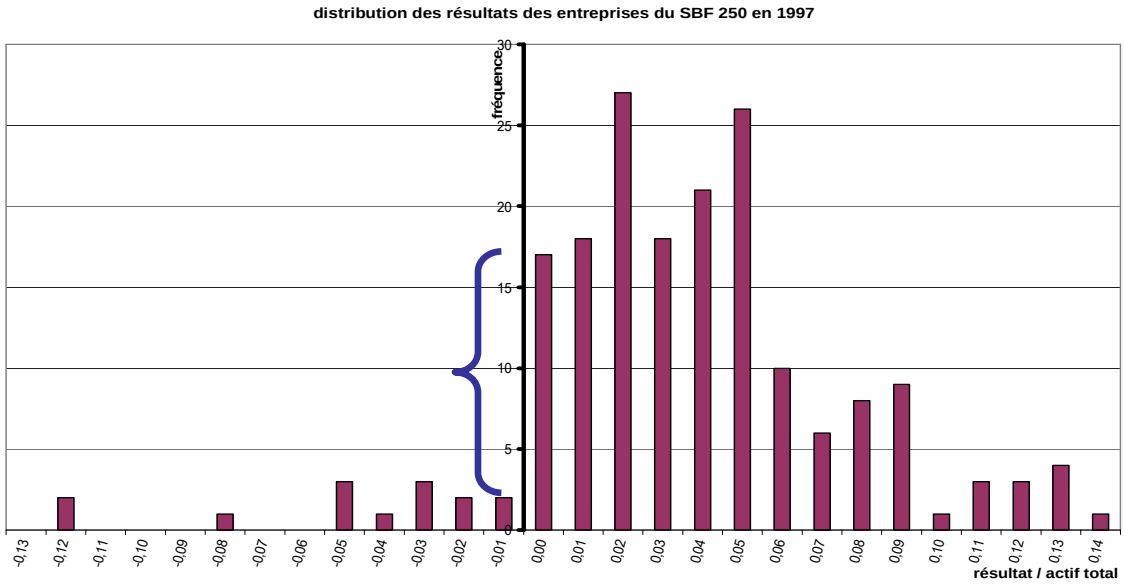
- Des mesures qui s'affranchissent de toute estimation théorique. Elles se contentent de comparer les effectifs observés entre eux. Ce sont des indicateurs d'asymétrie.

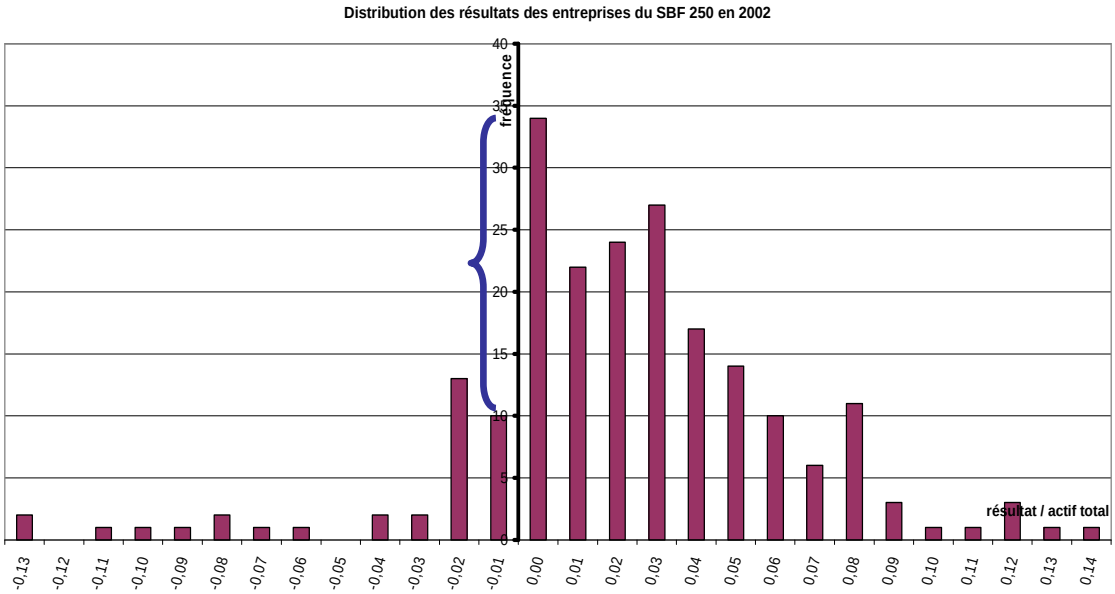
Les sous-parties b. et c. développent ces deux familles. Mais auparavant, la sous-partie a. évoque brièvement la méthode la plus intuitive d'identification des irrégularités : l'observation graphique.

a. Observation graphique des irrégularités de distribution

L'observation visuelle des irrégularités de distribution est assez simple à effectuer graphiquement. Les deux exemples ci-dessous montrent de manière évidente que l'effectif des entreprises faiblement bénéficiaires est chaque année sur évalué, et que l'effectif des entreprises faiblement déficitaires est toujours sous évalué.

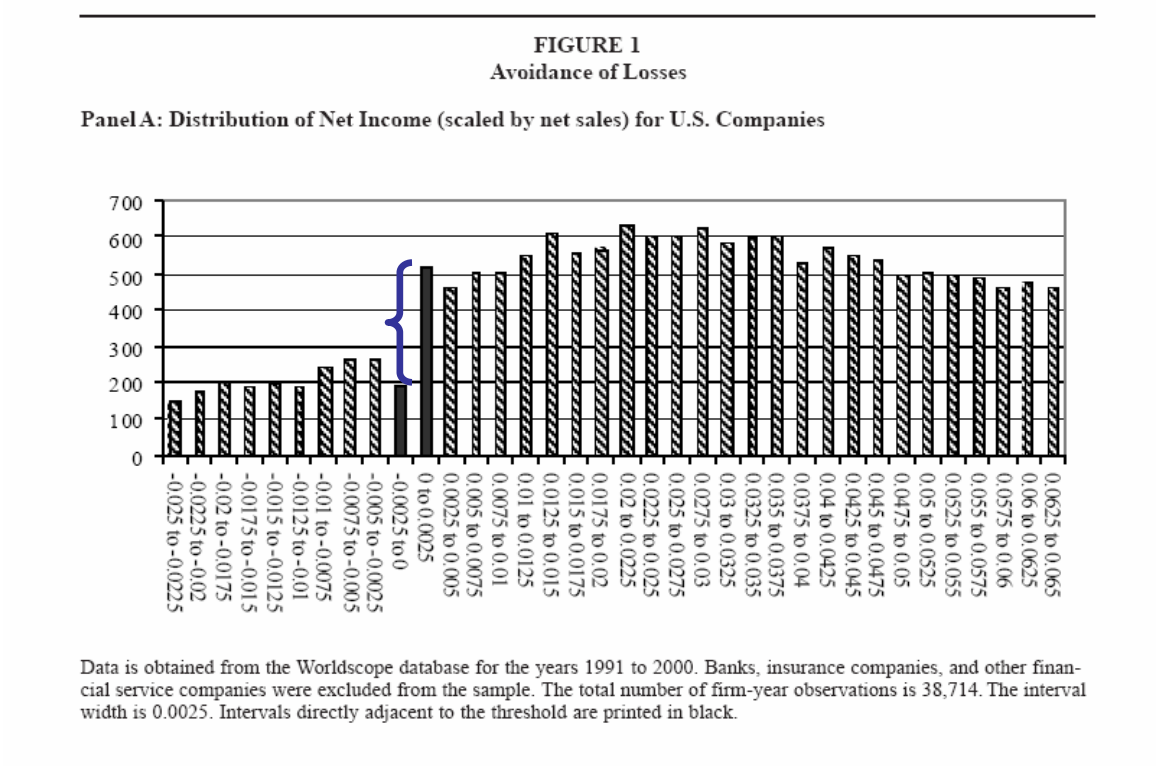
Figure 17 : Distribution du résultat mis à l'échelle par l'Actif Total de (n-1) des entreprises du SBF250 en 1997 et 2002





Ces deux exemples sont réalisés sur des populations relativement peu nombreuses (250 observations par année). Malgré cela, il apparaît graphiquement que l'irrégularité autour du zéro est toujours très forte. Sur une population plus importante, et sur une plus longue période, l'observation graphique de l'irrégularité est encore plus évidente. Ci-dessous, la distribution observée par Glaum, Lichtblau et Lindemann (2004) sur 38 714 observations et sur 10 ans :

Figure 18 : Figure issue de l'article de Glaum, Lichtblau et Lindemann (2004). La population représente près de 4 000 entreprises américaines sur 10 ans, soit 38 714 observations



À plusieurs reprises, dans la littérature, des auteurs se contentent d'une observation graphique pour illustrer un propos. Cette observation peut parfois pallier le fait que les tests statistiques ou les mesures effectuées ne suffisent pas à justifier les écarts constatés visuellement.

Figure 19 : Figure issue de l'article de Burgstahler et Dichev, 1997

Les auteurs écrivent : «Une observation visuelle des distribution peut compenser (…) et, dans ce cas, confirme la prédiction. Du panel A au Panel C, on observe une augmentation aussi bien du nombre des entreprises manipulatrices que de la mesure dans laquelle les résultats manipulés affectent des intervalles au-delà des deux intervalles immédiatement adjacents au zéro»[40].

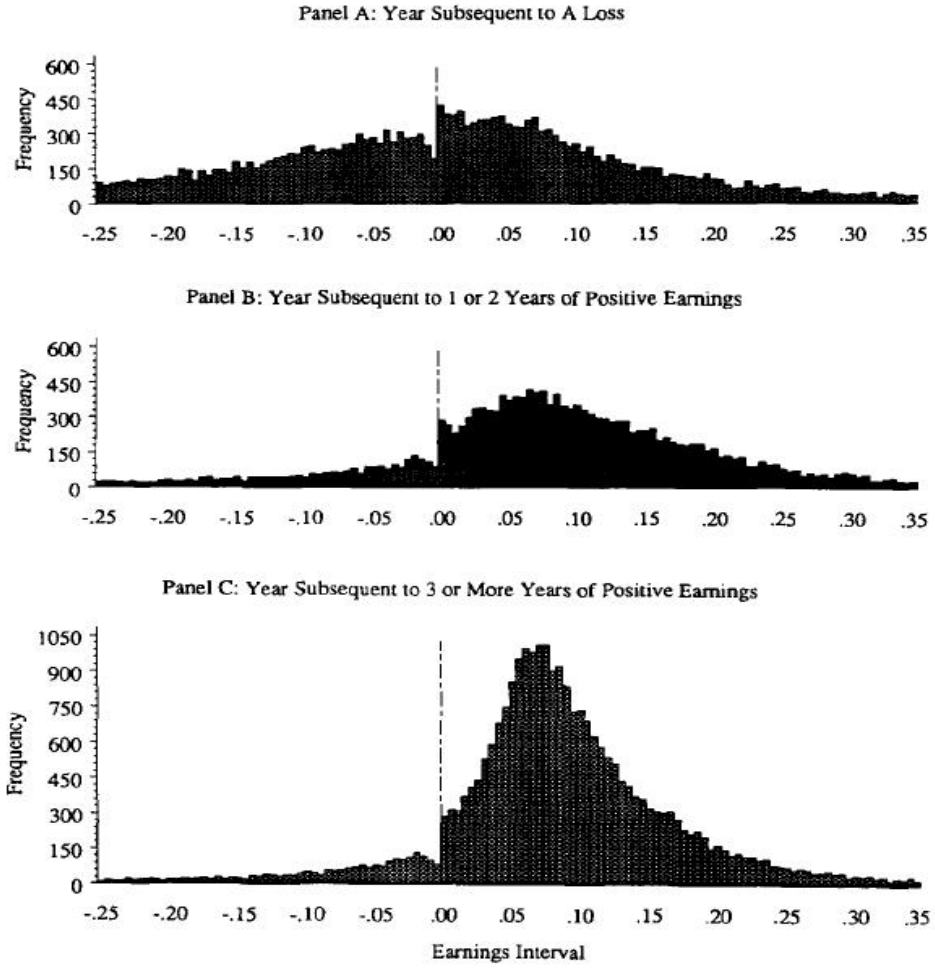


Fig. 4. Three empirical distributions of earnings scaled by market value categorized according to the pattern of preceding earnings for the firm. Panel A: the distribution for the years immediately following a loss; Panel B: the distribution for the years following exactly one or two years of positive earnings; and Panel C: the distribution for the years following three or more years of positive earnings. (See Fig. 3 for detailed definitions of variables.)

[40] Notre traduction de : « Visual inspection of the distributions can compensate (…) and, in this case, provides evidence consistent with the prediction. Moving from Panel A to Panel C, there is evidence of an increase in both the proportion of observations managed and the extent to which earnings management affects intervals other than the two intervals immediately adjacent to zero »

Figure 20 : Figure issue de l'article de Dechow, Richardson et Tuna, 2003

Les auteurs écrivent : « La figure (a) présente la distribution des résultats des entreprises cotées depuis moins de deux ans. Conformément avec l'hypothèse d'un biais de cotation, la discontinuité est plus prononcée pour ces entreprises récemment cotées. La figure (b) présente la distribution des résultats des entreprises qui ont plus de vingt ans. Une discontinuité plus petite est observable pour ces entreprises, mais elle demeure. Ces observations suggèrent que les conditions nécessaires pour entrer en bourse fournissent une explication partielle mais pas complète de l'existence des discontinuités. »[41]

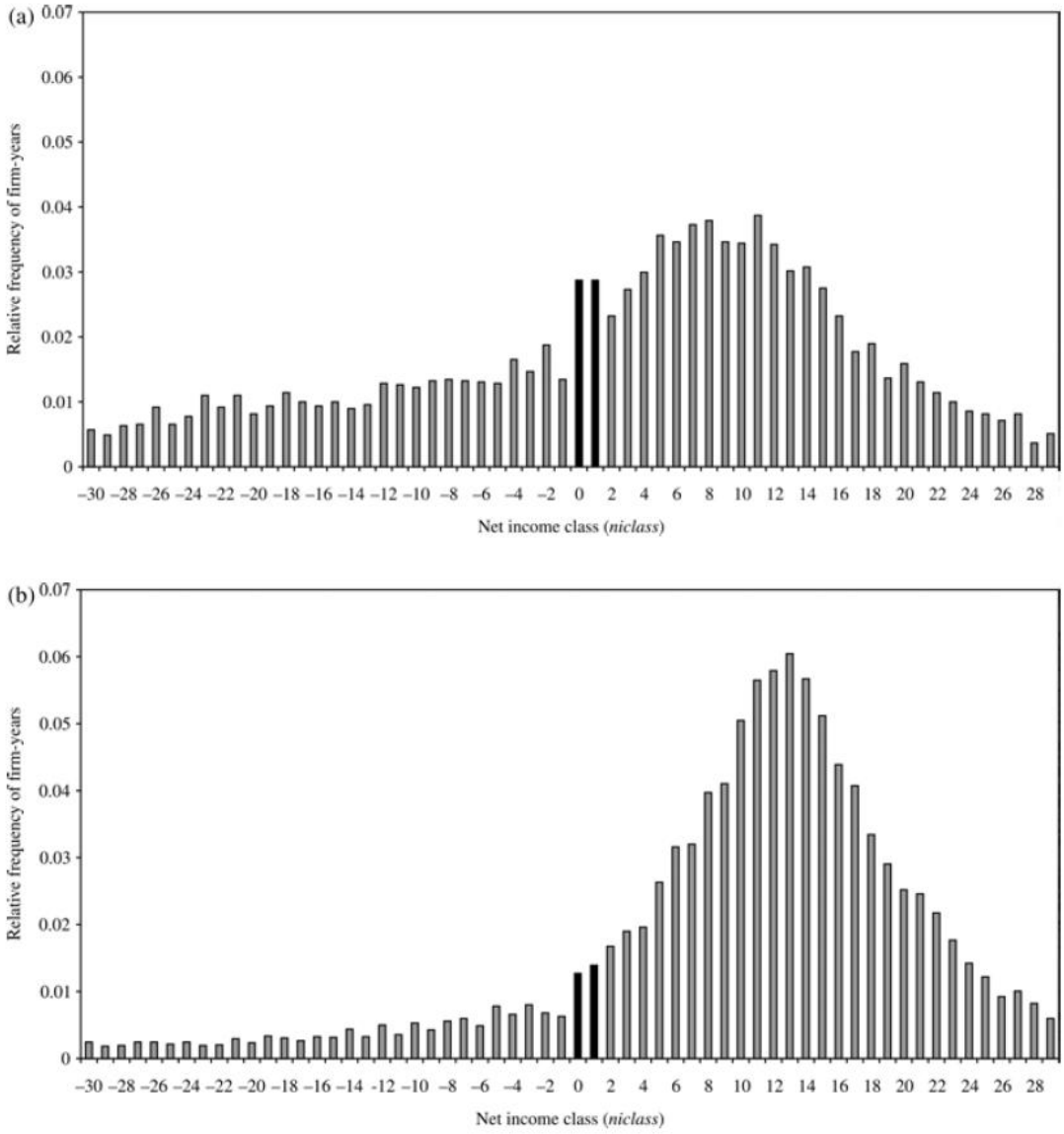


Figure 7. (a) The distribution of net income scaled by market value for firms listed for two years or less (4454 firm-years). (b) The distribution of net income scaled by market value for firms listed for more than twenty years (9016 firm-years).

[41] Notre traduction de : « Figure 7(a) provides a plot for firms that have been listed on an exchange for two years or less. Consistent with a listing bias, the kink is more extreme for newly listed firms. Figure 7(b) presents a plot of firms that are over twenty years of age. A smaller kink is observable for these firms, but the kink is still there. This suggests that exchange-listing requirements provide a partial but not complete explanation for the kink. »

Dans les deux cas ci-dessus (Figure 19 et Figure 20), de telles observations graphiques semblent cependant insuffisantes pour effectuer des comparaisons fiables. Elles relèvent davantage d'une intuition argumentée.

b. Mesures d'irrégularité

Pour s'assurer scientifiquement d'une irrégularité, les auteurs comparent l'effectif observé d'un intervalle de la population à un effectif théorique puis réalisent un test statistique. La question est alors de savoir comment estimer l'effectif théorique. Le problème est que personne ne connaît la loi de distribution théorique des résultats. Intuitivement, on considère que la distribution doit être plus ou moins gaussienne. Mais est-ce suffisant pour ajuster la distribution par une loi Normale ? Aucun auteur ne s'est encore aventuré à l'affirmer (ce point est développé dans la partie 3 de la thèse).

En fait, la littérature sur les seuils n'a pas arrêté une méthodologie homogène de mesure de l'effectif théorique, même si la plupart des auteurs s'inspirent de l'article fondateur de Burgstahler et Dichev (1997). Évitant la question de la loi théorique de distribution, les auteurs partent du postulat qu'en l'absence de manipulation, la distribution autour d'un intervalle quelconque devrait être lisse[42]. Ils estiment dès lors l'effectif d'un intervalle par comparaison avec les effectifs des intervalles adjacents[43]. Mard (2004) apporte un certain nombre de développements à cette méthode par moyenne arithmétique. Une méthode par symétrie est également utilisée par Burgstahler et Dichev (1997) et par Mard (2004). Dechow, Richardson et Tuna (2003) utilisent pour leur part des interpolations linéaire et exponentielle et Bisson, Dumontier et Janin (2004) ont recours à une extrapolation linéaire.

Globalement, ces démarches statistiques sont appelées « non paramétriques » car elles tentent de s'affranchir de l'étude des « paramètres » de la loi de distribution théorique des observations pour se limiter à l'étude des caractéristiques de la distribution aux abords du seuil étudié.

La description des différentes méthodes utilisées pour estimer l'effectif irrégulier s'articule ainsi : (1) les mesures linéaires (moyenne arithmétique, interpolation et extrapolation), (2)

[42] « Smooth » en anglais.

[43] Méthode non paramétrique d'estimation de l'effectif théorique.

les mesures tenant compte de la non linéarité de la courbe (symétrie, moyenne pondérée, interpolation exponentielle), (3) les mesures corrigée des transferts d'observations.

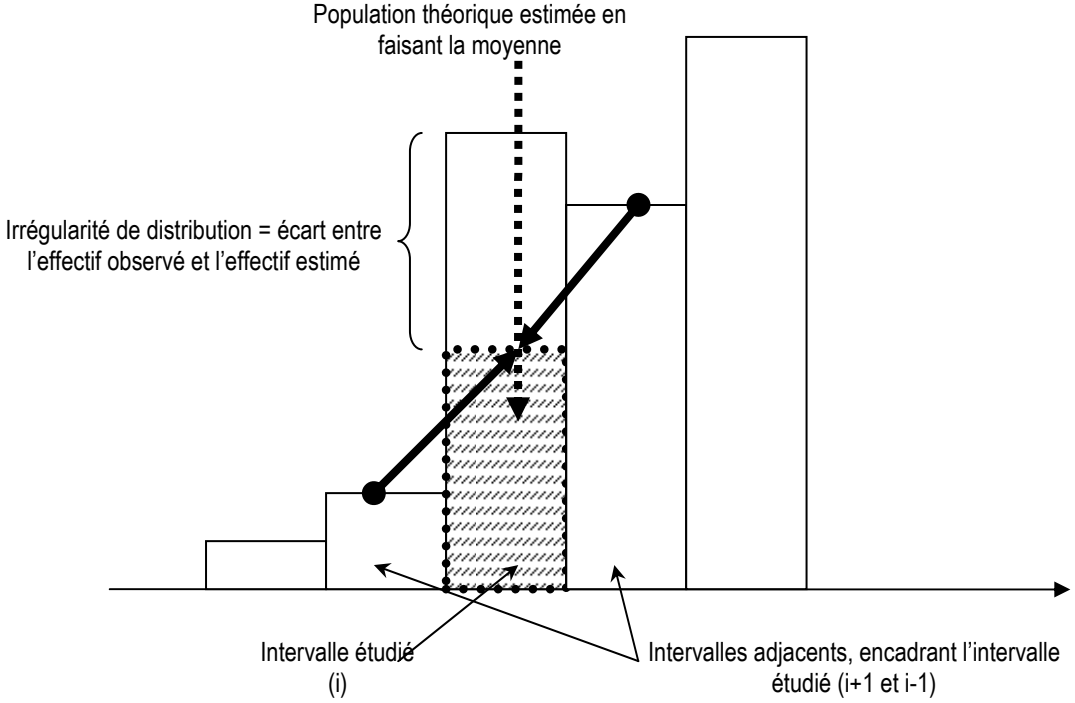
b.1- Mesures linéaires de l'amplitude

Les moyennes arithmétiques

Dans leur article fondateur, Burgstahler et Dichev (1997) font deux estimations successives : (1) « existence de la gestion du résultat »[44] est une estimation destinée à tester la significativité de l'irrégularité, et (2) « importance de la gestion du résultat »[45] est une estimation de l'ampleur[46] des manipulations. La seconde estimation fondée sur la symétrie supposée de la distribution est étudiée dans la sous-partie b.2.

La première mesure (« existence ») effectuée par les auteurs estime la population d'un intervalle i par la moyenne entre les populations des deux intervalles adjacents, c'est-à-dire qui encadrent l'intervalle étudié : i+1 et i-1. C'est une moyenne arithmétique.

Figure 21 : Mesure par moyenne arithmétique des deux intervalles adjacents



[44] En anglais, les auteurs parlent de « existence of earnings management ».

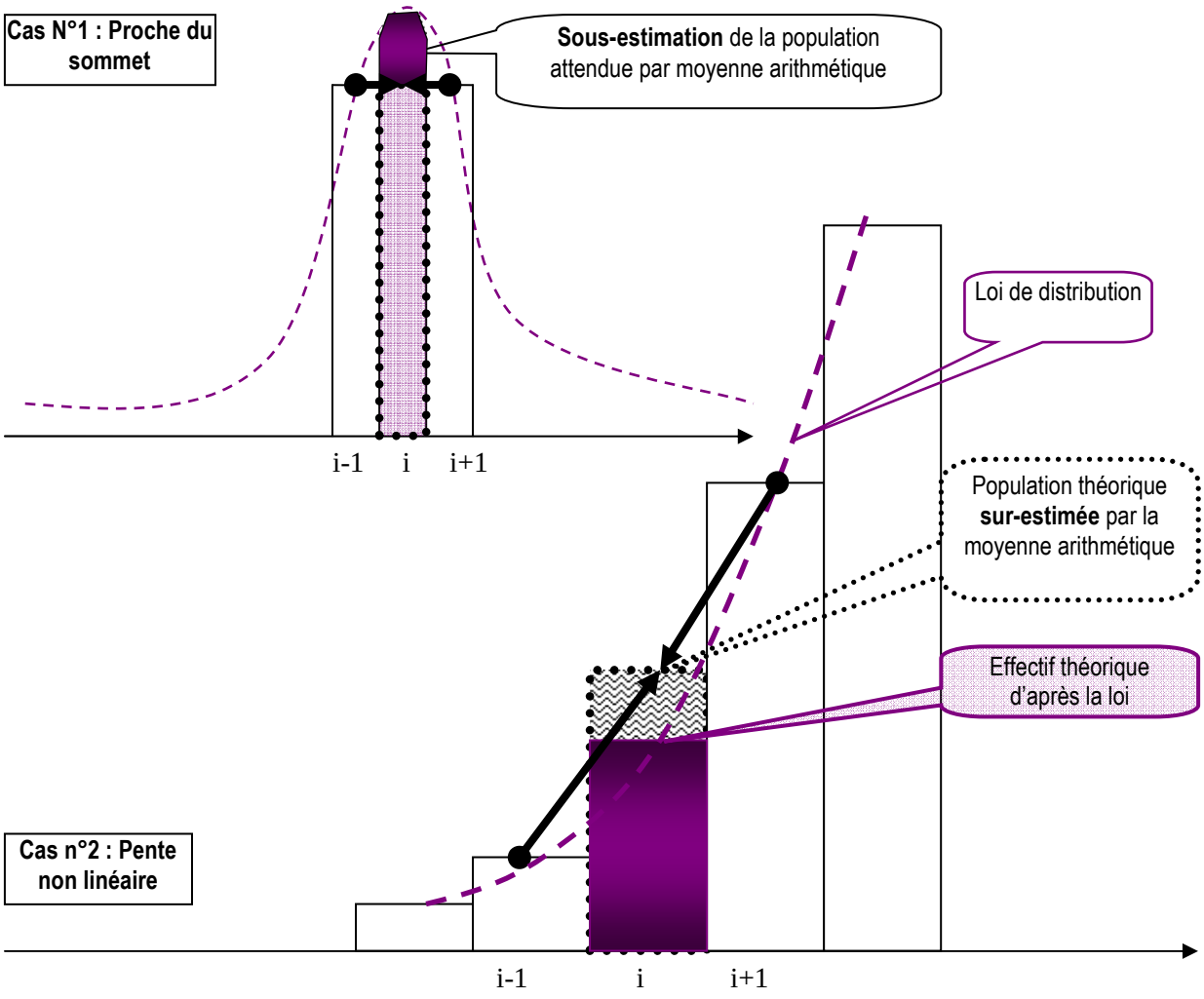
[45] En anglais, les auteurs parlent de « prevalence of earnings management ».

[46] Plus exactement, c'est une mesure de la « hauteur » de l'irrégularité, c'est-à-dire du nombre d'entreprises manipulatrices.

Cette approximation a un énorme avantage souligné par Burgstahler et Dichev eux-mêmes : il n'est pas utile de connaître la loi de distribution de la variable pour calculer l'effectif théorique. C'est une méthode qui privilégie la robustesse et la simplicité à la précision. L'estimation de l'effectif théorique par une moyenne arithmétique pose en effet trois problèmes majeurs :

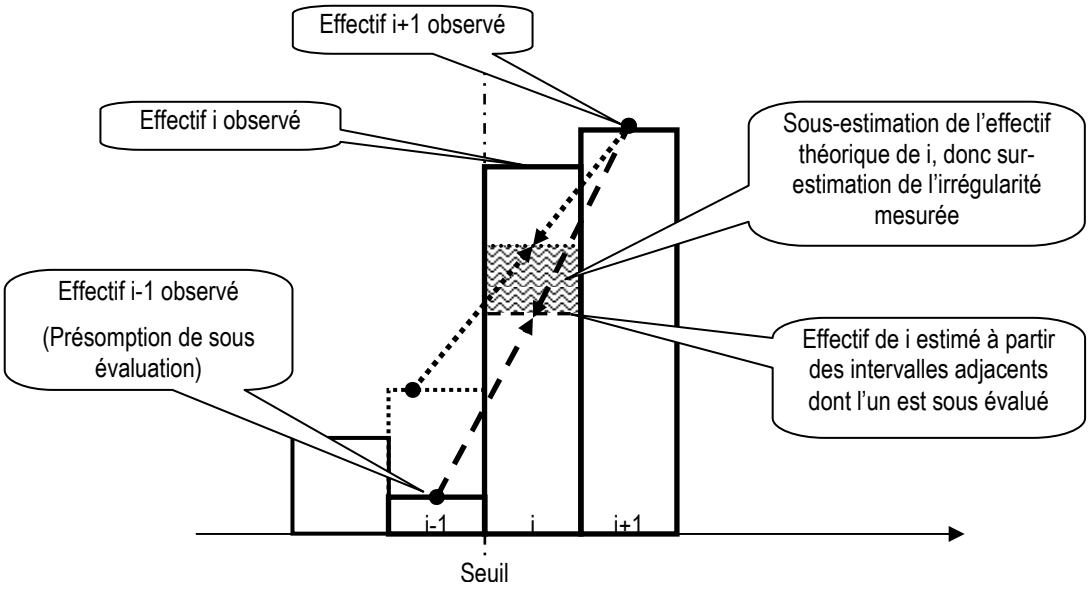
(1) Premièrement, lorsque l'effectif étudié est proche du sommet de la distribution, il y a inversion de la pente de la courbe. La moyenne entre les deux intervalles adjacents ne peut plus être une bonne approximation de la population étudiée. De manière plus générale, puisque la pente n'est pas une constante dans une distribution gaussienne (la distribution n'est ni linéaire, ni monotone), le calcul aura tendance à surestimer l'effectif lorsque l'on se situe sur une portion théorique de distribution concave (dérivée seconde supérieure à zéro), et à la sous-estimer sur des portions convexes (proche du sommet).

Figure 22 : Moyenne arithmétique, non monotonie et non linéarité de la distribution



(2) Deuxièmement, pour estimer la population d'un intervalle par la moyenne entre les deux intervalles adjacents, les auteurs font comme si les effectifs des deux intervalles adjacents sont fiables. Or les auteurs observent autour des seuils deux phénomènes distincts : une sur-représentation des entreprises situées au dessus du seuil, et une sous-représentation des entreprises situées en dessous du seuil. Il en découle que les mesures par moyenne arithmétique estiment un effectif théorique en interpolant des effectifs eux-mêmes présumés irréguliers. Ainsi, l'effectif théorique de l'intervalle situé juste au dessus du seuil aura tendance à être sous-estimé s'il est estimé à partir de l'intervalle juste inférieur (puisque lui-même sous-représenté). Ceci conduit en général à surestimer le nombre irrégulier d'entreprises faiblement bénéficiaires.

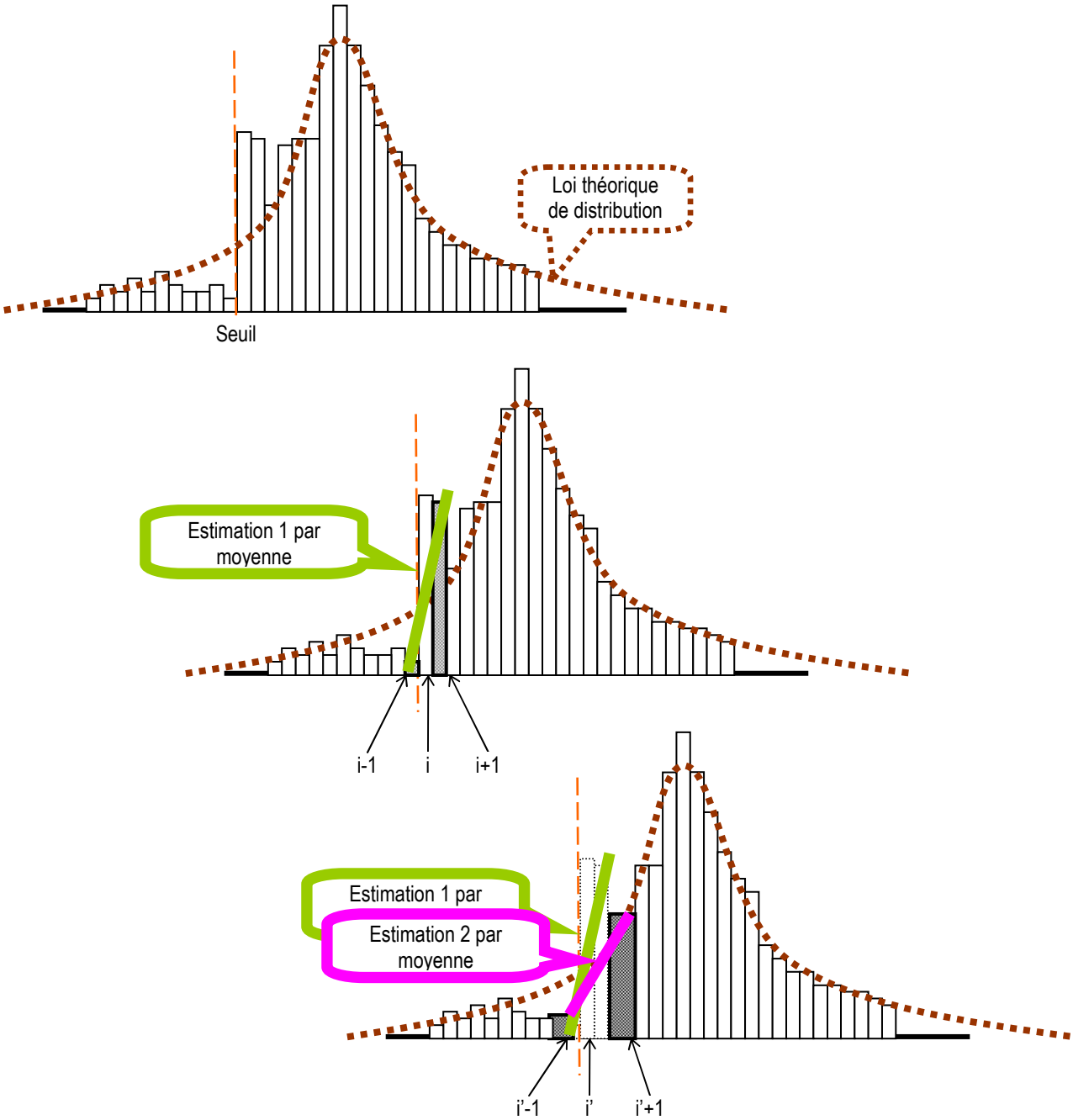
Figure 23 : Utilisation d'un intervalle potentiellement irrégulier pour mesurer l'irrégularité



Burgstahler et Dichev (1997) évoquent ce problème et tentent d'y répondre par deux tests : l'un pour l'intervalle juste supérieur, et l'autre pour l'intervalle juste inférieur à zéro. Cette précision ne supprime cependant pas le problème.

(3) Troisièmement, le choix de la largeur des intervalles, ou du nombre d'intervalles étudiés, peut influencer considérablement la mesure de l'irrégularité. C'est ce qu'illustre la Figure 24. L'estimation 1 de l'effectif théorique de l'intervalle i est réalisée en faisant la moyenne des intervalles i-1 et i+1. L'estimation 2 est réalisée en élargissant les intervalles. L'intervalle i' étudié est désormais égal à la somme des intervalles i et i+1. L'estimation est réalisée en calculant la moyenne des effectifs des intervalles i'-1 qui est la somme des intervalles i-1 et i-2, et i'+1 qui est la somme des intervalles i+2 et i+3.

Figure 24 : Influence de la largeur de l'intervalle d'étude sur les mesures par moyenne arithmétique



On constate ainsi graphiquement que modifier la largeur de l'intervalle d'étude peut modifier de manière très substantielle les estimations théoriques de l'effectif étudié, et en conséquence, la mesure des irrégularités.

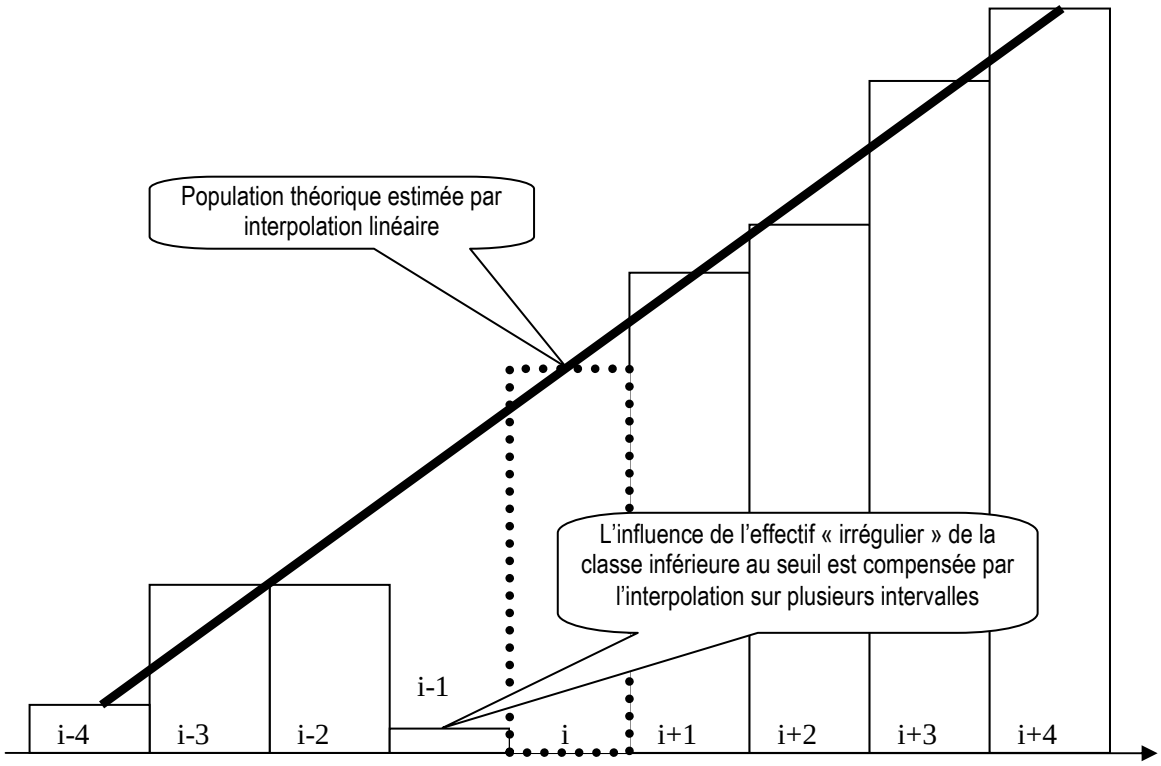
Malgré ces limites méthodologiques préoccupantes, les irrégularités sont suffisamment prononcées pour que les tests statistiques confirment l'existence d'une irrégularité

significative. En revanche, de telles mesures sont très insuffisantes pour effectuer des comparaisons.

Les interpolations linéaires

Dans leur article, Dechow, Richardson et Tuna (2003) estiment le nombre attendu d'entreprises dans un intervalle en utilisant deux méthodes : une interpolation linéaire et une interpolation exponentielle. L'interpolation exponentielle est étudiée dans la sous-partie b.2. Dans le cas de l'interpolation linéaire, les auteurs suivent une méthodologie proche de la moyenne arithmétique, mais en décalant leurs observations de 4 intervalles. Autrement dit, la pente est estimée entre les intervalles i-4 à i+4[47].

Figure 25 : Mesure par interpolation linéaire



Cette méthode répond partiellement aux critiques formulées à l'encontre de la méthode par moyenne. Elle résout le problème d'estimation d'un intervalle par un intervalle adjacent lui-même potentiellement irrégulier, à condition toutefois que la largeur des intervalles soit suffisante pour que l'intervalle inférieur (i-4) ne soit pas sous-estimé (comme pouvait l'être l'intervalle i-1). Cependant, elle ne permet pas de prendre en

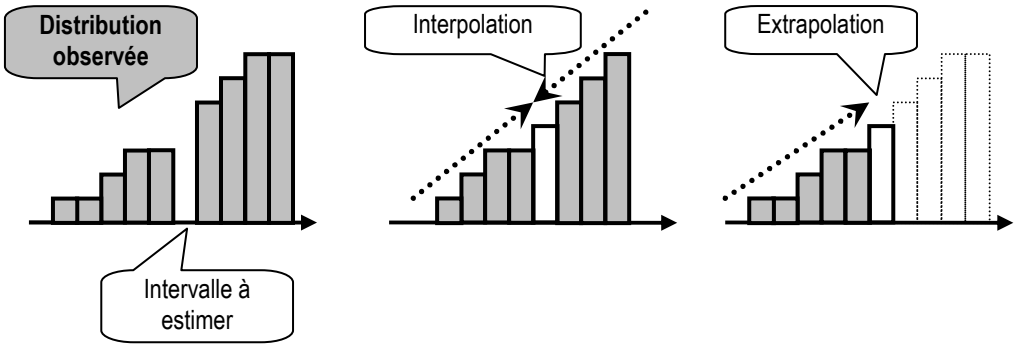
[47] Les auteurs estiment que 10 à 15% seulement d'entreprises sont irrégulièrement faiblement bénéficiaires, contre 30% selon le calcul par symétrie de Burgstahler et Dichev.

compte la pente non linéaire de la distribution et peut même accentuer cette erreur de calcul. Par ailleurs, elle reste très sensible à la largeur des intervalles.

Les extrapolations linéaires

Un seul papier de recherche (Bisson, Dumontier et Janin, 2004) a mis en œuvre une démarche par extrapolation. Une interpolation se fait lorsque l'on veut estimer un effectif inconnu entouré « de droite et de gauche » par des effectifs connus. On parle d'extrapolation lorsque l'on connaît une série d'effectifs et que l'on veut estimer un effectif « à droite » ou « à gauche » de la série connue.

Figure 26 : Mesure par extrapolation



Les calculs par extrapolation ne se justifient que si l'irrégularité est située à proximité du sommet de la distribution puisqu'en ce point, le renversement de la pente ne permet pas une interpolation.

Synthèse sur les calculs linéaires

En conclusion, les principaux avantages des calculs linéaires (moyenne arithmétique, interpolation ou extrapolation) méritent d'être soulignés : ils sont simples, et relativement robustes dans l'optique de mettre en évidence une irrégularité manifeste. Toutefois, ils posent les problèmes suivants :

(1) La largeur des intervalles (ou leur nombre) a une influence forte sur les résultats.

(2) La mesure se fait sous une hypothèse de linéarité de la courbe (pente constante) qui n'est pas justifiée. En pratique, les mesures atténuent les irrégularités situées sur la partie concave de la courbe.

(3) La mesure de l'effectif théorique d'un intervalle irrégulier par moyenne arithmétique se fait par rapport à ses voisins proches, potentiellement (et même sûrement) irréguliers eux-mêmes. La méthode de l'interpolation linéaire atténue ce problème, mais au prix d'une aggravation de la première limite.

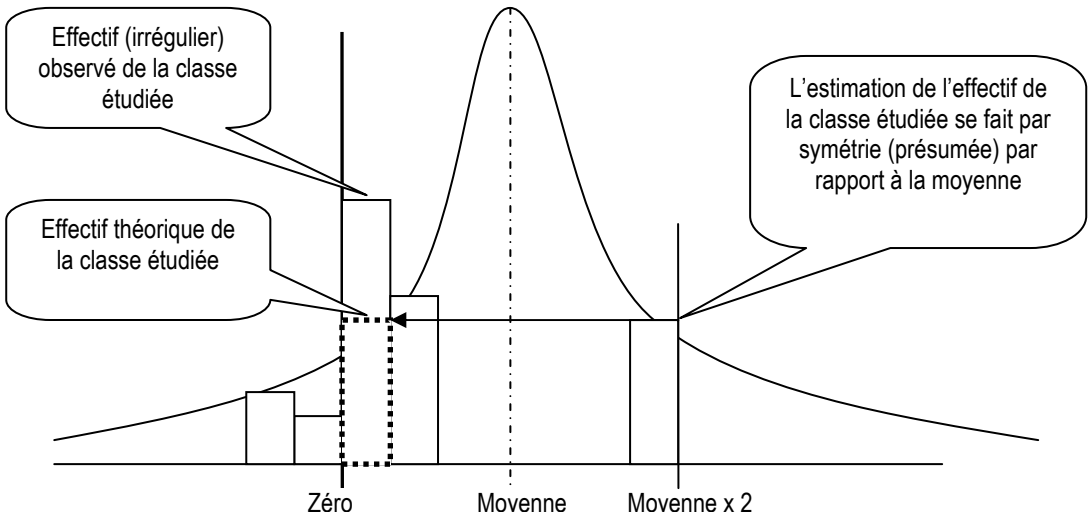
b.2- Mesure tenant compte de la non linéarité de la courbe

Après avoir mesuré « l'existence » par moyenne arithmétique, Burgstahler et Dichev (1997) effectuent une seconde estimation qu'ils appellent « prevalence[48] » afin d'évaluer le nombre d'entreprises ayant manipulé leurs comptes pour éviter le seuil. Pour effectuer cette mesure, ils postulent que la distribution des résultats doit être symétrique en absence de gestion du résultat.

Les mesures par symétrie

Burgstahler et Dichev (1997) estiment l'effectif de l'intervalle [-0,01 ; 0[[49] en prenant l'effectif de l'intervalle situé symétriquement à cet intervalle par rapport à la moyenne de la population. Ils en concluent que le nombre d'entreprises qui gèrent leur résultat est de l'ordre de 30 à 40 % de l'effectif situé sur le seuil.

Figure 27 : Mesure par symétrie



Par ailleurs, si Mard (2004) estime la population théorique dans le cas du seuil « résultat nul » par une moyenne (arithmétique, puis pondérée, puis corrigée), il utilise la symétrie (supposée) de la courbe dans le cas du seuil « variation nulle de résultat ».

[48] En français, « prevalence » se traduit par « prédominance ».

[49] Puis les auteurs font de même avec [-0,02 ; 0[et [-0,03 ; 0[pour contrôler la cohérence des résultats.

Dans les deux cas, les auteurs adoptent, dans le même article, deux méthodes différentes pour estimer un effectif théorique. Burgstahler et Dichev (1997) réalisent le test à partir d'une valeur théorique, puis calculent l'effectif théorique à partir d'une autre. Les auteurs estiment en effet qu'ils n'ont pas la certitude que les intervalles adjacents ne soient pas eux-mêmes irréguliers. Mard (2004) quant à lui justifie sa démarche par le fait que le second effet est situé beaucoup plus près du sommet de la distribution. L'auteur cherche à atténuer l'effet du renversement de la pente de la courbe aux alentours de la moyenne. Dans les deux cas, ces changements méthodologiques sont suscités par les limites techniques des mesures linéaires utilisées.

Il est intéressant de souligner que dans leurs articles respectifs, Mard (2004) comme Burgstahler et Dichev (1997) constatent la non symétrie de la courbe. Par exemple, Mard (2004) calcule un coefficient de symétrie (Skewness[50]) de -0,633 et un coefficient de concentration (Kurtosis[51]) de 3,913. Il évoque l'influence des valeurs extrêmes, mais la seule observation graphique de la distribution laisse penser que concentration comme dissymétrie ne sont pas uniquement liées à la présence de valeurs extrêmes[52].

En définitive, la démarche par symétrie est sujette à d'importantes limites méthodologiques puisqu'elle mélange des démarches paramétriques (postulat de symétrie de la loi de distribution) et non paramétriques (estimation ponctuelle de l'effectif d'un intervalle). Enfin, la mesure ne s'affranchit pas des problèmes de largeur des intervalles.

Les moyennes pondérées

Pour tenir compte de la non linéarité de la courbe, Mard (2004) estime l'effectif théorique à partir d'une moyenne pondérée des effectifs des intervalles voisins. La pondération cherche à prend en compte l'allure (supposée) gaussienne de la distribution.

Cette démarche soulève une question de principe : si les effectifs des intervalles adjacents sont estimés par une loi normale, pourquoi continuer à estimer l'effectif de

[50] Le Le coefficient d'asymétrie (Skewness) est une mesure du degré de symétrie d'une courbe. S'il est nul, la courbe est parfaitement symétrique autour d'une valeur moyenne. S'il est positif, on parle d'asymétrie positive (la pente est plus douce à droite qu'à gauche). S'il est négatif, on parle d'asymétrie négative (la pente est plus douce à gauche qu'à droite). Un Skewness de -0,6 correspond à une distribution asymétrique dont la queue de distribution est étalée vers la gauche.

[51] Le Kurtosis est un coefficient d'aplatissement. Égal à zéro, on considère que la courbe est plate. Le Kurtosis d'une loi Normale est égal à 3. Si le Kurtosis est supérieur à 3, on peut considérer la distribution comme significativement leptokurtique (c'est-à-dire « anormalement » pointue). Inférieur à 3, on considère la distribution comme platykurtique (aplatie).

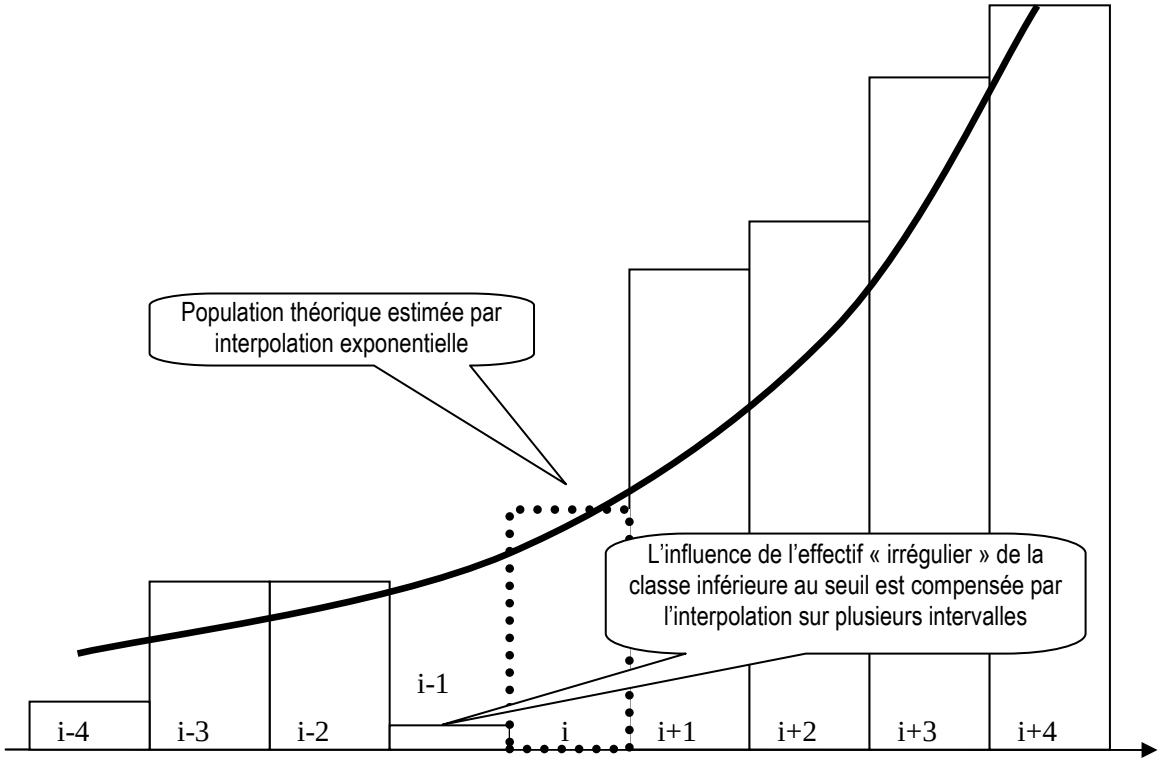
[52] Ce point est développé dans la partie 3 de la thèse.

l'intervalle étudié par un calcul de moyenne ? Pourquoi ne pas estimer directement l'effectif par la loi normale ? Par ailleurs, le postulat de Normalité est discutable dans la mesure où l'auteur constate la dissymétrie et la concentration de la courbe. Enfin, cette démarche ne résout pas deux des problèmes déjà énoncés : la largeur des intervalles a une influence forte sur les résultats, et l'estimation de l'effectif théorique d'un intervalle se fait par rapport à des intervalles eux-mêmes potentiellement irréguliers.

Les interpolations exponentielles

Dechow, Richardson et Tuna (2003) apportent une réponse originale et simple au problème de la non linéarité de la courbe de distribution. Ils estiment l'effectif théorique par interpolation exponentielle (en plus d'une interpolation linéaire déjà décrite dans la sous-partie b.1.). La pente étant estimée entre les intervalles i-4 et i+4, leur solution résout tout à la fois le problème de la non linéarité et le problème de l'utilisation d'intervalles eux-mêmes irrégulières pour estimer l'effectif théorique.

Figure 28 : Mesure par interpolation exponentielle



Ce calcul est cependant soumis à plusieurs limites : (1) il est toujours sensible à la largeur des intervalles retenue, et (2) il suppose une allure exponentielle de la courbe de distribution des résultats. Cette dernière hypothèse est sans doute adaptée lorsque l'on

se situe sur une portion de la courbe concave (phase d'accélération de la pente de la courbe). Mais elle ne l'est pas lorsque la courbe est convexe, c'est-à-dire proche du sommet, ni sur les queues de distribution. (3) Elle est enfin totalement inadaptée sur la moitié droite de la courbe[53].

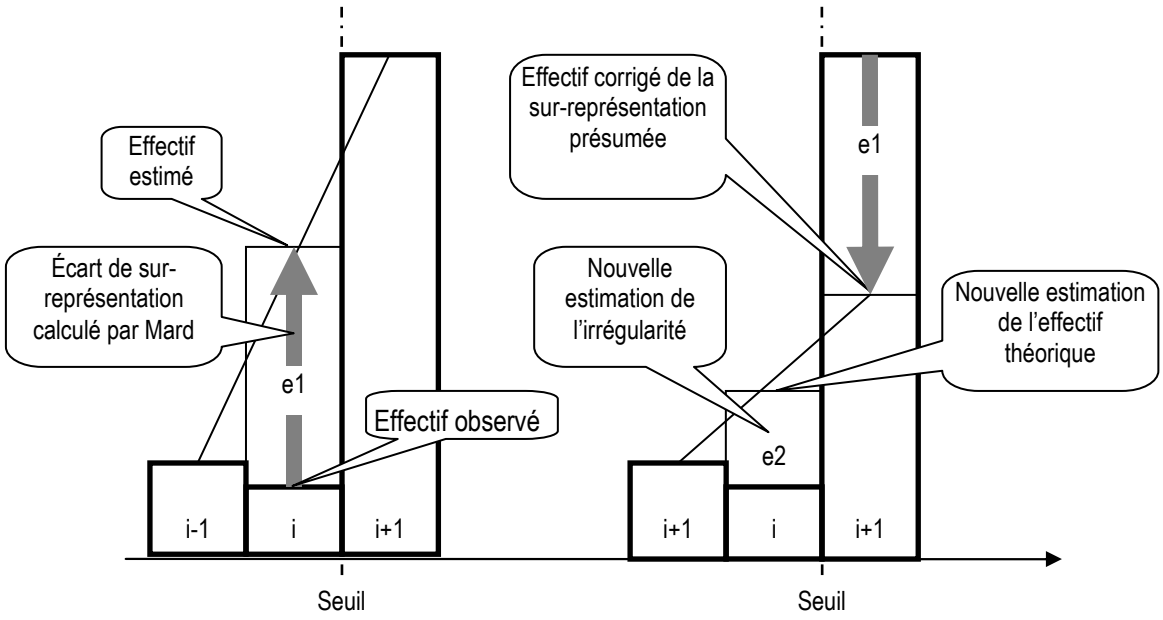
Les interpolations logarithmiques

Il est remarquable qu'aucun auteur ne fasse référence à une interpolation logarithmique. Pourtant, en se rapprochant du sommet de la courbe, la pente d'une distribution gaussienne diminue, puis s'annule (avant de s'inverser) et une telle interpolation pourrait se justifier. Ce point sera abordé dans l'étude empirique (section A du chapitre VI).

b.3- Mesures tenant compte des transferts d'observations

Le calcul se fonde sur le postulat que l'intervalle juste supérieur au seuil est surestimé, alors que l'intervalle juste inférieur est sous-représenté. Les calculs de moyenne estiment un de ces intervalles en utilisant le second, ce qui biaise les calculs. Mard (2004) propose donc une mesure corrigée de ces irrégularités liées aux transferts (présumés) d'effectifs d'un intervalle vers l'autre. La moyenne est calculée en diminuant de l'intervalle juste supérieur à l'intervalle étudié l'excédent calculé sur l'intervalle étudié.

Figure 29 : Prise en compte des transferts d'observations

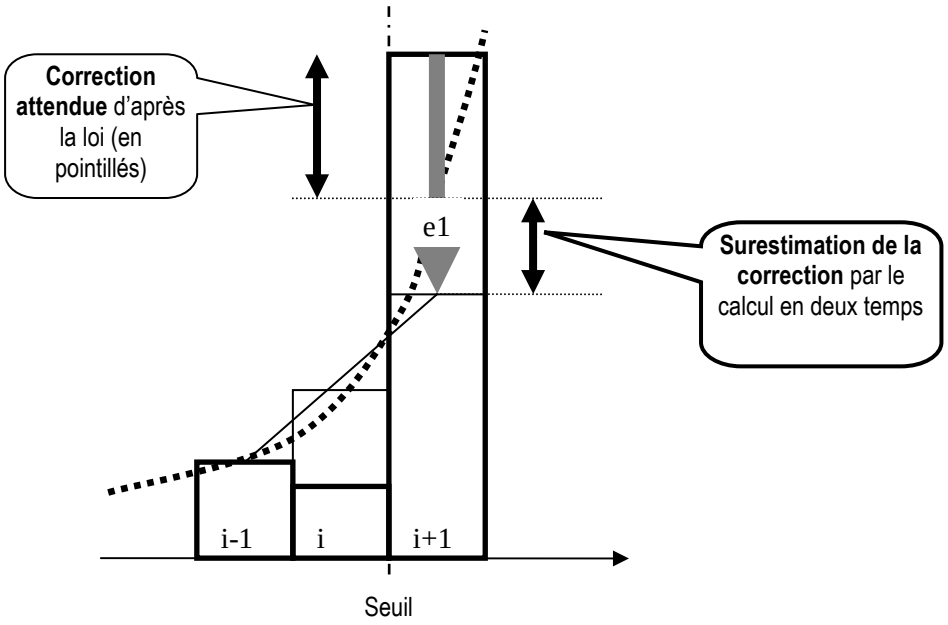


[53] Pour que le seuil du résultat nul soit situé sur la moitié droite de la courbe, il faut que la moyenne des résultats annuels des entreprises soit négative. Ce cas se présente rarement, mais il intervient deux fois entre 1992 et 2004 (en 2002 et 2003) dans l'échantillon d'entreprises françaises étudié dans le cadre des applications empiriques de la thèse.

Ce correctif astucieux n'en demeure pas moins approximatif puisque le calcul se fait en deux temps. C'est ce qu'illustre la Figure 29. Dans un premier temps, l'effectif de l'intervalle i étudié est estimé par moyenne arithmétique. On en déduit un écart e1. Dans un deuxième temps, on re-estime l'effectif de l'intervalle i en diminuant l'effectif de l'intervalle i+1 de la valeur calculée de e1. On obtient alors une nouvelle mesure plus atténuée de l'irrégularité (e2).

Ce calcul en deux temps est approximatif puisque cette nouvelle estimation de l'irrégularité (e2) devrait conduire à une révision de la correction de l'effectif de l'intervalle i+1. Autrement dit, i+1 qui était présumé sur-estimé devient désormais sous-estimé. L'irrégularité doit en effet se situer quelque part entre e1 et e2. Il serait ainsi possible de faire plusieurs itérations du calcul avant d'aboutir à une estimation fiable. Un calcul de limite de suite serait sans doute plus adapté. La Figure 30 illustre par un exemple graphique le fait que la correction peut être largement surestimée. Sur cette figure, on peut comparer visuellement le correctif (e1) effectué au correctif qu'il aurait fallu prendre en compte si la loi de distribution était connue (courbe en pointillés).

Figure 30 : Limite du calcul en deux temps de prise en compte des transferts d'observations



Par ailleurs, ces mesures, demeurent très sensibles à la largeur de l'intervalle retenu. Enfin, ce correctif suppose que le transfert se fait en totalité de l'intervalle juste inférieur au seuil vers l'intervalle juste supérieur selon l'hypothèse (présentée dans la revue de littérature de la partie 1) de manipulation marginale pour éviter le seuil. Or cette

hypothèse n'a jamais été testée dans la littérature, et les résultats de la partie 3 de la thèse tendent à la réfuter.

c. Les indicateurs d'asymétrie

Plusieurs études internationales ou longitudinales se fondent sur une mesure d'asymétrie pour effectuer des comparaisons de seuils. Il est important de distinguer irrégularité et asymétrie :

- L'irrégularité est l'écart entre l'effectif observé et un effectif de référence (ou effectif attendu). C'est la « sous » ou « sur » représentation.
- L'asymétrie est la comparaison des effectifs de deux intervalles successifs, dont l'un au moins est présumé être irrégulier.

Les études qui se fondent sur des mesures d'asymétrie évitent le problème de la mesure des irrégularités. Mais elles reposent sur une hypothèse non vérifiée : la sous-représentation à gauche, et la sur-représentation à droite d'un seuil sont deux effets liés à un même phénomène d'évitement. Cette hypothèse rejoint le postulat de manipulation marginale pour dépasser un seuil déjà présenté dans la revue de littérature de la partie 1. L'asymétrie entre les deux intervalles mesure-t-elle deux phénomènes dépendants ? Y a-t-il une corrélation entre l'effectif à droite du seuil et celui à gauche ? Cette hypothèse n'a jamais été testée.

Conceptuellement, le terme même d' « asymétrie » laisse supposer que la valeur de référence est la « symétrie ». Autrement dit, la neutralité supposerait l'égalité des effectifs sur les deux intervalles consécutifs. Glaum, Lichtblau et Lindemann (2004) disent ainsi que leur indicateur est centré sur zéro. En réalité, la distribution suivant une forme (plus ou moins) de cloche, il n'y a jamais deux intervalles consécutifs égaux (sauf éventuellement au sommet). Constater une asymétrie n'est donc pas nécessairement synonyme d'irrégularité. Il apparaît dès à présent que les indicateurs d'asymétrie posent des problèmes d'interprétation.

Deux indicateurs d'asymétrie sont recensés dans la littérature. Ils sont présentés successivement.

c.1. L'indicateur d'asymétrie « d/g » (ou « r/l »)

Le premier indicateur d'asymétrie utilisé dans la littérature est le ratio « effectif faibles gains » divisé par « effectif faibles pertes » (Leuz, Nanda et Wysocki, 2003 ; Burgstahler, Hail et Leuz, 2006).

n_r = effectif de l'intervalle de droite (r = right)

n_l = effectif de l'intervalle de gauche (l = left)

L'indicateur d'asymétrie est mesuré par le ratio = n_r / n_l

Ce ratio est souvent appelé par simplification « r/l » ou, en français, « d/g ».

C'est un calcul extrêmement simple, repris par Dechow, Richardson et Tuna (2003) pour étudier l'évolution dans le temps (de 1989 à 2001) des « pics ». Les auteurs calculent ce ratio en réunissant deux intervalles (0 et 1 sur -1 et -2)[54]. Si la mesure d'asymétrie d/g est simple à calculer, elle soulève plusieurs limites mathématiques :

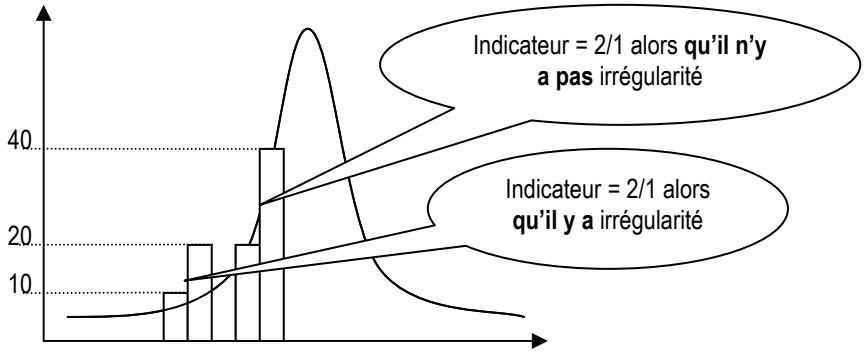
(1) Pour des petits nombres (population réduite, ou largeur des intervalles faible), l'effectif à gauche de l'irrégularité peut être égal à zéro. Le dénominateur est alors nul et l'indicateur devient inopérant.

(2) L'indicateur est compris entre 0 et l'infini. Si les deux effectifs sont égaux, il est égal à un. La distribution de l'indicateur est donc asymétrique et non linéaire, rendant difficile les comparaisons et les interprétations. Par exemple, la différence entre une asymétrie de 1,5 et 2 est nettement plus importante qu'une différence entre une asymétrie de 10 et 10,5. Il est par ailleurs impossible de comparer une irrégularité située sur la partie droite de la courbe à une irrégularité située sur la partie gauche (par exemple une irrégularité de valeur 2/1 est identique à une irrégularité de valeur ½).

(3) Par ailleurs, on ne peut pas considérer que l'indicateur soit centré sur 1 puisque cette situation n'est valable que si la courbe est plate (donc au sommet). L'indicateur ne tient donc absolument pas compte de l'endroit où l'on se situe dans la courbe (supposée en cloche). Un ratio de 2 peut être tout à fait normal lorsque l'on se trouve dans la partie concave de la courbe, alors que c'est une irrégularité importante lorsque l'on est dans la partie convexe (proche du sommet), ou sur les queues de distribution.

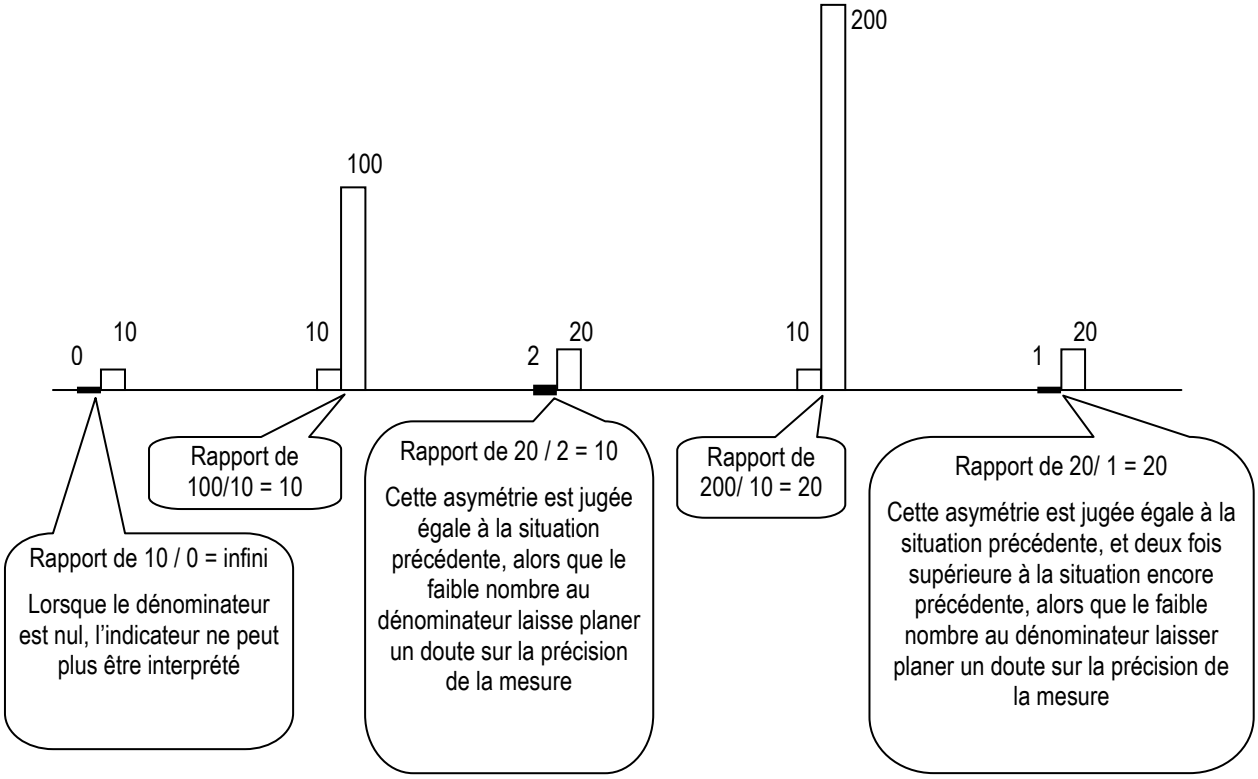
[54] Les auteurs utilisent une largeur d'intervalle de 0,01, donc le ratio est mesuré sur une largeur de 0,02.

Figure 31 : Indicateur d'asymétrie d/g et non linéarité de la distribution



(4) L'indicateur est par ailleurs très sensible à la taille de l'échantillon, en exagérant les valeurs extrêmes.

Figure 32 : Indicateur d'asymétrie et taille de l'échantillon



En conclusion, comme le soulignent Glaum, Lichtblau et Lindemann (2004), l'indicateur d/g est assez rudimentaire et est mal approprié pour effectuer des comparaisons d'irrégularités dans le contexte de la gestion du résultat.

c.2. L'indicateur « (r-l)/(r+l) »

Glaum, Lichtblau et Lindemann (2004) ont créé un indicateur de mesure de l'asymétrie qui évite un certain nombre des « propriétés numériques indésirables[55] » de l'indicateur de Leuz, Nanda et Wysocki (2003). Les auteurs s'inspirent d'un indicateur largement utilisé dans les sciences de la nature (Halzen et Martin, 1984), et appellent leur indicateur A comme « Asymétrie ».

$A = (n_r - n_l) / n_{rl}$

Avec :

n_r = effectif de l'intervalle de droite (r = right)

n_l = effectif de l'intervalle de gauche (l = left)

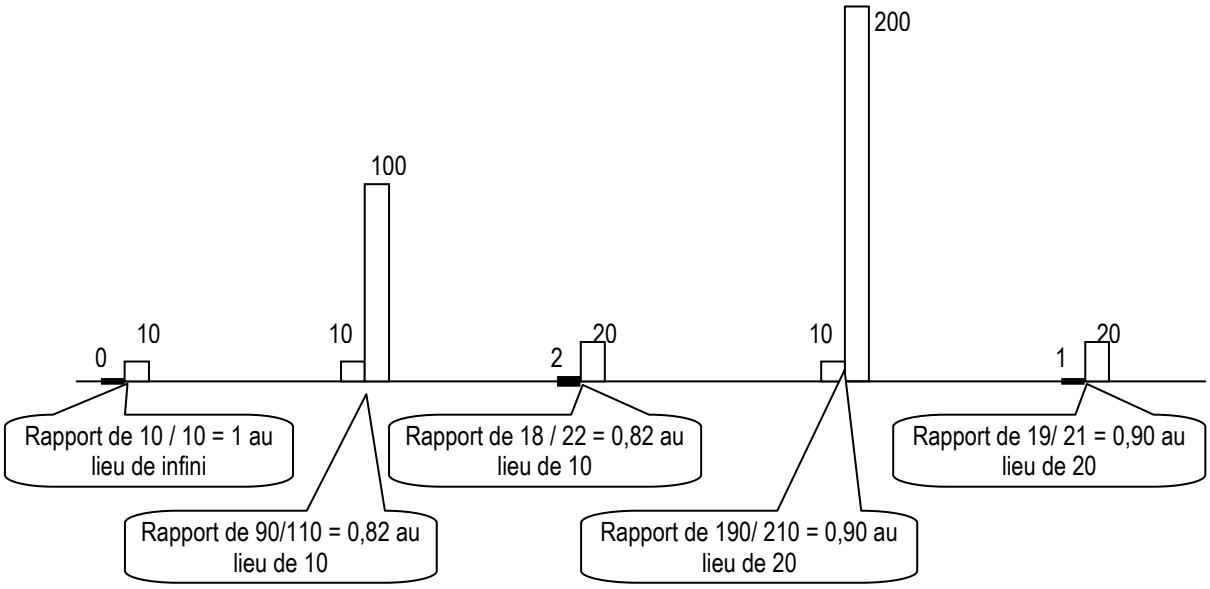
n_{rl} = somme des effectifs des deux intervalles

En simplifiant l'expression, A = (r-l)/(r+l) et en la francisant, on obtient A = (d-g)/(d+g).

Contrairement à l'indicateur de Leuz, Nanda et Wysocki (2003), A est centré sur 0, et compris entre -1 et 1. La distribution de l'indicateur est donc (en apparence) symétrique rendant a priori plus facile les comparaisons. Pour des petits nombres, un effectif à gauche du seuil égal à zéro (donc un dénominateur nul) ne rend pas l'indicateur inopérant. Il est par ailleurs moins sensible à la taille de l'échantillon puisque le dénominateur est moins « volatile ».

[55] Selon l'expression des auteurs.

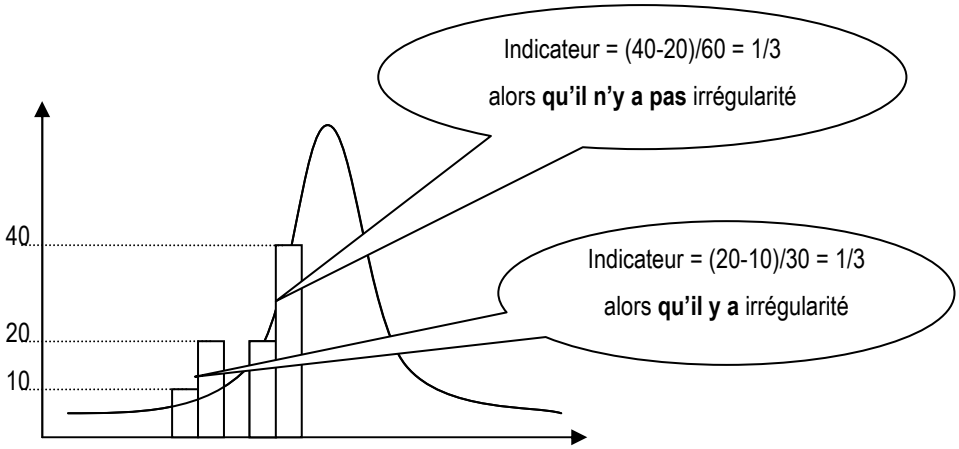
Figure 33 : Comparaison des deux indicateurs d'asymétrie



Comme le montre la Figure 33 ci-dessus, les deux indicateurs fournissent la même hiérarchie des asymétries, mais les valeurs extrêmes sont lissées avec le second indicateur.

Cet indicateur présente cependant des limites. Tout d'abord, il n'est toujours pas « linéaire ». L'écart entre une asymétrie de 0,1 et 0,2 n'est pas comparable à l'écart entre une asymétrie de 0,8 et 0,9. Mais le plus préoccupant, c'est que l'indicateur ne prend toujours pas en considération la non linéarité de la courbe. En réalité, la distribution de l'indicateur ne doit pas être considérée comme réellement symétrique et centrée sur 0 puisque cette situation n'est valable que si la courbe est plate (donc au sommet).

Figure 34 : Indicateur d'asymétrie (r-l)/(r+l) et non linéarité de la distribution



La Figure 34 illustre par exemple le fait qu'un score de 0,33 peut être tout à fait normal lorsque l'on se trouve dans la partie croissante de la courbe, alors que c'est une irrégularité importante lorsque l'on est dans la partie haute, ou basse de la courbe.

c.3. Synthèse sur les indicateurs d'asymétrie

Le principal problème posé par l'utilisation d'indicateurs d'asymétrie est un problème conceptuel. Ces indicateurs considèrent en effet que l'effet de seuil se définit par la somme des deux effets : sur-représentation à droite, et sous-représentation à gauche. Or aucune étude n'a jamais vérifié la dépendance de ces deux effets. Cette hypothèse mérite d'être testée à partir d'une mesure d'irrégularité séparant les deux effets. Certains des résultats de Glaum, Lichtblau et Lindemann (2004) peuvent s'expliquer par cette limite méthodologique (cet aspect a déjà été discuté dans la revue de littérature de la partie 1 de la thèse, lors de la discussion de la Figure 13).

D'autres limites peuvent être énumérées :

- L'indicateur d'asymétrie ne s'affranchit pas des problèmes de largeur de l'intervalle déjà discutés dans le cadre des mesures d'irrégularité non paramétriques. Ce problème est d'autant plus important que l'indicateur est utilisé pour comparer des populations de tailles différentes, et pour lesquelles les largeurs d'intervalles retenues peuvent être différentes.

- Les mesures sont faites sur la totalité des effectifs d'un intervalle, et non pas sur les effectifs irréguliers.

- La mesure de l'asymétrie ne tient pas compte de l'endroit où l'on se trouve sur la courbe ; il ne répond pas au problème de non linéarité de la pente de la courbe de distribution, rendant toute interprétation de l'indicateur délicate.

d. Synthèse des limites des mesures non paramétriques

Tableau 5 : Mesures non paramétriques des irrégularités. Postulats implicites et synthèse des limites et avantages

Postulat implicite sur la distribution	Méthode	Avantages	Limites	Auteurs
Linéarité	Moyenne arithmétique	Simplicité des calculs	Ne tient pas compte de la non linéarité de la pente ; Sensible à la largeur des intervalles; Se fonde sur un intervalle irrégulier pour estimer un intervalle théorique	Burgstahler et Dichev, 1997 ; Mard, 2004
	Interpolation linéaire	Limite l'impact de la largeur des intervalles	Ne tient pas compte de la non linéarité de la pente (et accentue ses conséquences) ;	Dechow, Richardson et Tuna, 2003
	Extrapolation linéaire	Est adaptée lorsque l'irrégularité est très proche du sommet	Ne tient pas compte de la non linéarité de la pente	-
Symétrie	Symétrie	Tient compte de la non linéarité	Se fonde sur un postulat de symétrie non vérifié ; Est impossible à calculer lorsque la discontinuité est proche du sommet ; Est sensible à la largeur des intervalles	Burgstahler et Dichev, 1997 ; Mard, 2004
Non linéarité	Interpolation exponentielle	Est adaptée en phase d'accélération de la pente ; Limite l'impact de la largeur des intervalles	Est inadaptée lorsque la pente décélère (au sommet) et sur les queues de distribution	Dechow, Richardson et Tuna, 2003
	Interpolation logarithmique	Est adaptée sur la portion proche du sommet de la distribution ; Limite l'impact de la largeur des intervalles	Est inadaptée lorsque l'irrégularité s'éloigne du sommet de la distribution	-
	Extrapolation exponentielle	Sont adaptées lorsque l'irrégularité est très proche du sommet	Se fondent sur un postulat de distribution non vérifié	-
	Extrapolation logarithmique			
Transfert d'observations	Moyenne pondérée	Tient compte de la non linéarité	Se fonde sur un postulat de distribution normale non vérifié ; Est sensible à la largeur des intervalles ; Se fonde sur un intervalle irrégulier pour estimer un intervalle théorique	Mard, 2004
	Moyenne pondérée corrigée	Tient compte de la non linéarité	Le calcul en deux temps peut largement surestimer la correction ; Est sensible à la largeur des intervalles ; Se fonde sur un intervalle irrégulier pour estimer un intervalle théorique	Mard, 2004

Tableau 6 : Synthèse des limites et avantages des indicateurs d'asymétrie

Postulat implicite sur la distribution	Méthode	Avantages	Limites	Auteurs
Les indicateurs d'asymétrie n'estiment pas un effectif théorique. Aucun postulat n'est fait implicitement sur la distribution.	d/g	Simplicité des calculs	Ne tient pas compte de la non linéarité de la pente ; Sensible à la largeur des intervalles; Se fonde sur un intervalle irrégulier pour estimer un intervalle théorique	Dechow, Richardson et Tuna, 2003 ; Leuz, Nanda et Wysocki, 2003 ; Burgstahler, Hail et Leuz, 2006
	(d-g)/(d+g)	Simplicité des calculs	Ne tient pas compte de la non linéarité de la pente (et accentue ses conséquences) ;	Glaum, Lichtblau et Lindemann, 2004

2. Démarche de l'étude méthodologique

Après avoir présenté les limites théoriques des différentes méthodes non paramétriques de mesure des irrégularités, une étude empirique est menée afin de mettre en évidence l'impact réel de la méthode de mesure adoptée sur la valeur obtenue de l'irrégularité.

a. Mise en évidence de l'influence de la méthode retenue

La démarche est la suivante : l'irrégularité au seuil du résultat nul est mesurée sur les distributions des résultats des entreprises françaises année après année, de 1992 à 2004. Pour cela, l'effectif théorique de la distribution est estimé selon les différentes méthodes répertoriées dans le Tableau 5. Chaque année, deux mesures sont effectuées : l'une mesure l'irrégularité à gauche du seuil, l'autre l'irrégularité à droite.

Chaque irrégularité est exprimée en pourcentage. L'effectif observé est divisé par l'effectif théorique. Cet indicateur relatif est extrêmement simple à calculer et à interpréter. Il permet de comparer des seuils dans le temps ou dans l'espace. L'irrégularité relative a l'énorme qualité d'être intuitivement interprétable[56]. Par exemple, une irrégularité égale à

[56] La mesure de la significativité de l'irrégularité centrée réduite (la différence standardisée) est utilisée par Brown et Caylor (2005) comme indicateur de comparaison. Cet indicateur dépend de la valeur de l'irrégularité, et de l'écart type de la différence, lui-même dépendant de la taille de la population et de la probabilité qu'une entreprise se situe sur l'intervalle étudié (donc indirectement de la largeur et du nombre d'intervalles). Le principal défaut de cet

0,5 (ou 50 %) est immédiatement compréhensible puisqu'elle indique que l'effectif observé est moitié moindre que ce qui est attendu.

Les différentes mesures sont ensuite comparées. L'étude est motivée par l'hypothèse que le choix d'une méthode influence sensiblement les résultats des mesures. Cette hypothèse doit donc être confirmée.

b. Hiérarchisation des méthodes non paramétriques de mesure

Après avoir mis en évidence le fait que le choix d'une méthode de mesure n'est pas neutre sur les résultats obtenus, l'étude cherche à déterminer quelle est la méthode la plus fiable dans le souci de fournir des résultats concrets et utilisables dans de futures recherches.

Pour y parvenir, cinq critères de qualité sont étudiés : la volatilité, la qualité de la régression, l'analyse de classification, la comparaison des tendances, et la sélection d'un indicateur fondée sur l'emplacement de l'irrégularité.

b.1. La volatilité des mesures

Plus la volatilité d'une mesure est grande, plus la méthode est sensible aux choix de mesure que sont les valeurs extrêmes, la largeur des intervalles, ou les aléas de la distribution qui n'est jamais totalement lisse même en absence de manipulation. Une volatilité élevée doit donc être interprétée comme un signe de fragilité de la méthode.

b.2. La qualité de la régression

Un critère de classement des méthodes d'estimation réside dans sa capacité à bien représenter les effectifs observés. Autrement dit, les méthodes nécessitant une régression (interpolation et extrapolation) peuvent être classées en fonction du R^2 de la fonction utilisée pour prédire les effectifs.

b.3. L'analyse de classification hiérarchique

Le dendrogramme des mesures d'irrégularités peut être tracé de manière à faire apparaître des familles de méthodes. Sans préjuger de la qualité de la mesure, le

indicateur est que son interprétation n'est pas intuitive, c'est pourquoi une mesure relative de l'irrégularité est préférable.

dendrogramme permet d'identifier les méthodes dont les mesures convergent des méthodes qui fournissent des résultats atypiques.

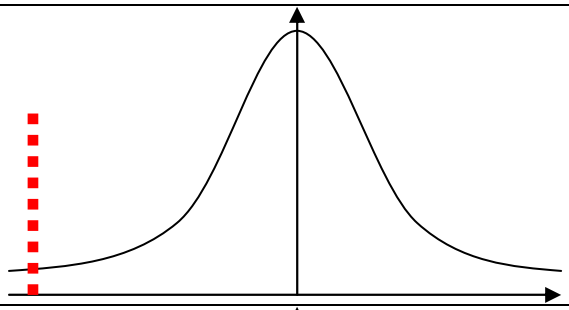
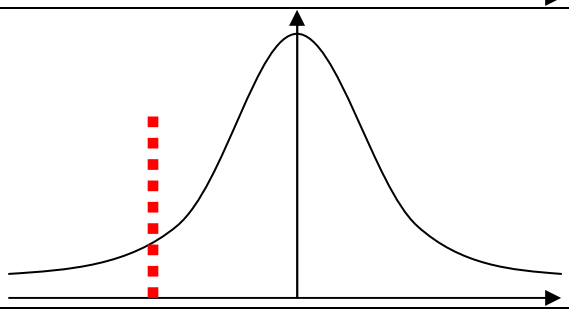
b.4. La comparaison des tendances

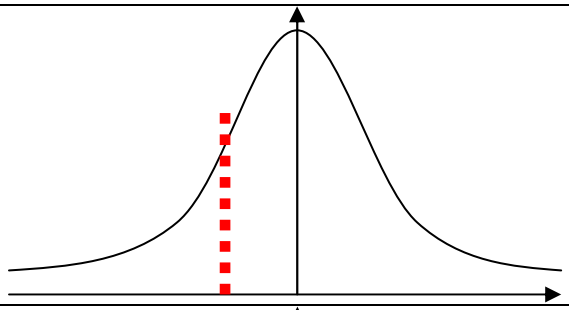
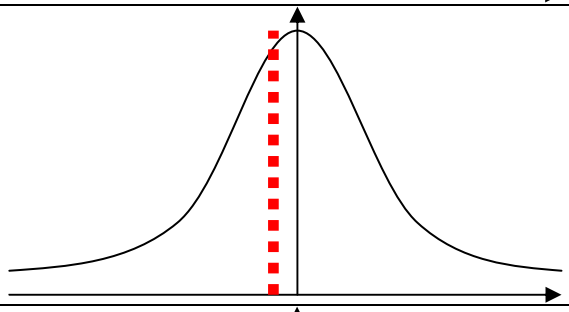
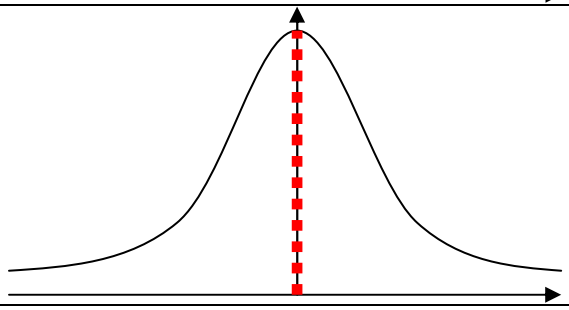
Chaque irrégularité étant mesurée sur 13 années selon chaque méthode, chaque méthode conduit à constater une évolution temporelle différente. Comme dans le cas de la classification hiérarchique, comparer les tendances ainsi dégagées ne permet pas de juger la qualité des méthodes, mais permet d'identifier les méthodes atypiques des méthodes convergentes.

b.5. La pertinence de l'indicateur en fonction de l'emplacement de l'irrégularité dans la distribution

Si l'on considère que la distribution attendue n'est pas linéaire, mais doit avoir une forme plus ou moins en cloche, il est alors théoriquement possible de proposer la méthode la plus pertinente en fonction de l'emplacement de l'irrégularité dans la distribution. La Figure 35 récapitule ces choix.

Figure 35 : Choix de la mesure appropriée en fonction de la position du seuil sur la distribution

Situation du seuil sur la distribution	Représentation graphique	Caractéristique de la distribution aux alentours du seuil	Mesure appropriée
Queue de distribution		Pente douce	Interpolation linéaire
Partie concave de la distribution		Pente en accélération (portion concave)	Interpolation exponentielle

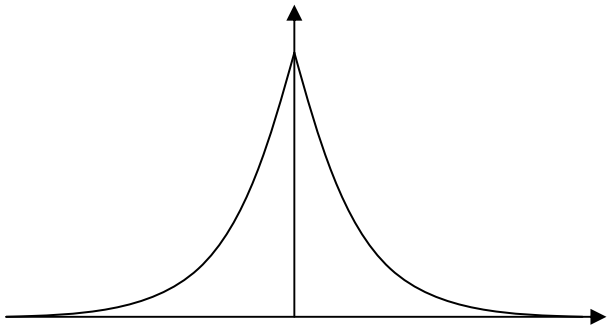
Partie croissante de la distribution		Pente en croissance stable	Interpolation linéaire
Proche du sommet		Pente positive en décélération (portion convexe)	Interpolation logarithmique
Sommet de la distribution		Inversion de la pente, en passant par une pente nulle	Une interpolation devient impossible[57] ; une extrapolation (linéaire ou logarithmique) est préférable

La Figure 35 ne s'intéresse pas aux cas où le seuil se situe sur la partie droite de la courbe (ce qui est plus rare mais pas impossible). Une réflexion similaire peut être menée en transformant préalablement les données (en multipliant par -1).

Il est important de souligner que les graphes de la Figure 35 représentent une allure en cloche très générale. La réflexion menée dans ce paragraphe n'implique pas que la distribution soit strictement Normale ou symétrique. Cette allure en cloche est cependant discutable. D'autres allures pourraient être proposées. Une allure doublement exponentielle (Figure 36) est par exemple envisageable.

[57] Elle conduit inévitablement à une sous-estimation de l'effectif estimé.

Figure 36 : Remise en cause de l'allure en cloche de la distribution



Si une telle allure devait être retenue, alors l'interpolation linéaire ou l'extrapolation logarithmique proposées lorsque l'irrégularité se rapproche du sommet (deux dernières lignes du tableau de la Figure 35) ne seraient pas justifiées. Fondamentalement, il apparaît que toute réflexion théorique pour déterminer la pertinence de la mesure ne peut faire l'économie d'une réflexion sur l'allure attendue de la distribution générale des résultats. Si l'étude empirique dans le cadre de la partie 2 sera menée en se référant à une allure « en cloche », la partie 3 quant à elle reviendra sur ce postulat.

c. Analyse de la pertinence des indicateurs d'asymétrie

Suite à l'étude des mesures non paramétriques des irrégularités, la recherche développe une deuxième série de travaux concernant les indicateurs d'asymétrie. Concrètement, cette étude mesure le degré d'indépendance des irrégularités à gauche et à droite du seuil du résultat nul. En effet, les auteurs utilisant des indicateurs d'asymétrie pour comparer les irrégularités mélangent les deux phénomènes. Ces indicateurs ne sont donc pertinents que si les deux mesures sont fortement corrélées.

Chacune des 10 années de l'échantillon, la valeur des deux discontinuités à droite et à gauche du seuil résultat nul est donc mesurée, et la corrélation entre ces deux valeurs est examinée.

d. Résultats attendus de l'étude empirique sur le choix de la méthode de mesure

L'étude sur l'impact du choix de la méthode de mesure de l'irrégularité doit répondre à trois questions : (1) Les différentes méthodes recensées dans la littérature fournissent-elles des résultats similaires, ou l'influence du choix est-il au contraire suffisamment

importante pour remettre en cause les conclusions des recherches ? (2) Si l'impact du choix de la mesure ne peut être négligé, alors quel(s) est (sont) l' (les) indicateur(s) qu'il faut privilégier ? (3) Les indicateurs d'asymétrie utilisés dans la littérature sont-ils des instruments de comparaison fiables des irrégularités ?

B. Impact de la largeur des intervalles d'observation

Pour mesurer une irrégularité, la distribution est divisée en intervalles[58]. Les effectifs théoriques et/ou réels sont alors estimés et/ou observés sur chaque intervalle.

1. Présentation du problème

La plupart des auteurs se réfèrent à des mesures théoriques de la largeur des intervalles, sans pour autant les appliquer formellement. Degeorge, Patel et Zeckhauser (1999) font référence à la largeur théorique $2(IQR)n^{-1/3}$ (Silverman, 1986 ; Scott, 1992) où IQR[59] est l'intervalle interquartile de la distribution et n le nombre d'observations. Dans une note de bas de page, Holland et Ramsey (2003) énumèrent cinq formules de calcul : $3{,}5\sigma n^{-1/3}$ (Scott, 1979), $2(IQR)n^{-1/3}$ (Freedman et Diaconis, 1981), $1{,}06\sigma n^{-1/5}$ (Silverman, 1986), $0{,}9(IQR)n^{-1/5}$ (Silverman, 1986), et $0{,}9(\min(\sigma, IQR/1.34))n^{-1/5}$ (Silverman, 1986). Les auteurs précisent que théoriquement, ces méthodes ne sont pertinentes que si la distribution est proche d'une distribution Normale.

À l'exception de Bisson, Dumontier et Janin (2004) et Van Caneghem (2006), dans la pratique, chaque auteur retient in fine une largeur ad hoc permettant une mesure de l'irrégularité. Ainsi, la largeur la plus fréquemment retenue est 0,005, c'est-à-dire une amplitude de résultat de l'ordre de 5 ‰ du total actif (Burgstahler et Dichev, 1997 ; Burgstahler et Eames, 2003 ; Dechow, Richardson et Tuna, 2003 ; Mard, 2004 ; Brown et Caylor, 2005 ; Coppens et Peek, 2005 ; Durtschi et Easton, 2005 ; Beaver, McNichols et Nelson, 2007) ou 0,01 (Leuz, Nanda et Wysocki, 2003 ; Phillips, Pincus et Rego, 2003) voire 0,0025 (Glaum, Lichtblau et Lindemann, 2004) sans justifier mathématiquement un tel choix.

[58] Ou « Classe » ou « pas ».

[59] En anglais = Interquartile Range.

Le paradoxe, c'est que, comme le montrent Glaum, Lichtblau et Lindemann (2004), selon la largeur choisie, les conclusions de certaines études peuvent être modifiées. Dans cette section de la thèse, l'étude du lien entre largeur des intervalles et irrégularité est approfondie dans le but d'améliorer la fiabilité des mesures effectuées.

2. Démarche de l'étude méthodologique

L'étude cherche à mettre en évidence l'impact concret de la largeur des intervalles sur la mesure de l'irrégularité.

a. Démarche

L'irrégularité est mesurée autour du seuil en faisant varier la valeur de la largeur de l'intervalle d'observation. Ces mesures sont effectuées en utilisant les deux méthodes[60] de mesure de l'irrégularité utilisées par Burgstahler et Dichev (1997) dans leur article fondateur (moyenne des deux intervalles adjacents, et symétrie). Une attention particulière est accordée aux largeurs ad hoc (0,005 ; 0,01) généralement utilisées dans la littérature, et à la largeur théorique ($2(IQR)n^{-1/3}$).

b. Résultats attendus de l'étude empirique sur l'impact de la largeur des intervalles

Les résultats attendus sont de 5 ordres :

(1) L'étude vise à confirmer le fait que la largeur des intervalles influence considérablement la mesure de l'irrégularité, et ce, quelle que soit la méthode de mesure retenue.

(2) L'étude vise à vérifier la pertinence du choix d'une largeur déterminée mathématiquement selon la formule théorique $2(IQR)n^{-1/3}$.

(3) L'étude cherche à identifier, pour chaque type de mesure, des contraintes techniques qui limitent l'éventail des largeurs possible à utiliser.

[60] L'utilisation de la dizaine de mesures identifiées dans la section A aurait été trop laborieuse. L'étude s'est donc focalisée sur les deux indicateurs « historiques ».

(4) L'étude débouche enfin sur une réflexion sur la notion de largeur optimale de l'irrégularité. Si la largeur influence la mesure, alors la largeur optimale est celle qui maximise l'irrégularité. Cette largeur optimale est alors proposée comme nouvel indicateur permettant de mieux appréhender le phénomène à l'origine des irrégularités de distribution de résultat.

En conclusion, l'attention est attirée sur le fait que les limites relatives à la largeur des intervalles sont induites par la nature non paramétrique des méthodes de mesure des irrégularités employées jusque-là par les chercheurs. Dans la mesure où les effectifs théoriques sont estimés à partir d'effectifs observés et que les effectifs ne sont jamais parfaitement lisses, toute mesure est inévitablement vouée à être influencée par la largeur de l'intervalle de référence. Cette réflexion conduit à envisager un instrument paramétrique de mesure indépendant de l'impact du choix de la largeur des intervalles.

C. Impact de la variable de mise à l'échelle

1. Présentation du problème

Les études sur les discontinuités des distributions de résultat utilisent toujours une variable de résultat « réduite » (ou mise à l'échelle) par un indicateur de taille (dénominateur) : Valeur de Marché (VM), Actif Total (AT) ou Chiffre d'Affaires (CA)[61].

a. La valeur de marché

La plupart des études (essentiellement anglo-saxonnes) utilisent la valeur de marché (VM) des capitaux propres pour mettre à l'échelle le résultat, la variation de résultat ou les prévisions (Burgstahler et Dichev, 1997 ; Kasznik et McNichols, 2002 ; Kinney, Burgstahler et Martin, 2002 ; Matsumoto, 2002 ; Burgstahler et Eames, 2003 ; Burgstahler, Elliott et Hanlon, 2003 ; Dechow, Richardson et Tuna, 2003 ; Holland et Ramsey, 2003 ; Phillips, Pincus et Rego, 2003 ; Payne et Thomas, 2004 ; Xue, 2004 ; Brown et Caylor, 2005 ; Shen et Chih, 2005 ; Burgstahler et Eames, 2006 ; Beaver,

[61] Par la suite, ce sont les expressions abrégées VM, AT et CA qui seront utilisées dans la thèse.

McNichols et Nelson, 2007). De même, Durtschi et Easton (2005) réduisent le résultat par action par la valeur de marché de l'action[62].

Ce choix de dénominateur est fortement remis en cause par Dechow, Richardson et Tuna (2003) qui soulignent que l'évaluation des entreprises faiblement déficitaires n'est pas faite selon la même méthode que celle des entreprises faiblement bénéficiaires. Les évaluations seraient plus orientées sur les valeurs liquidatives du bilan. C'est ce que confirment Durtschi et Easton (2005). Les auteurs en déduisent que mettre à l'échelle par la valeur de marché doit accentuer les mesures d'irrégularité.

Les auteurs utilisent $VM_{(n-1)}$, la valeur de marché du début de l'année, pour mettre le résultat à l'échelle. En effet, $VM_{(n-1)}$ est un indicateur de taille indépendant de l'activité de l'année, donc indépendant du résultat. Il est important que la variable de mise à l'échelle soit indépendante de la variable d'étude. Ce faisant, les études perdent mécaniquement la première année d'observation et augmentent très sensiblement le nombre de données manquantes puisque pour qu'une observation soit réalisée, il faut que les informations soient présentes deux années successives. Les entreprises qui entrent dans la cote (et donc dans la base de données) sont systématiquement exclues la première année. L'analyse empirique menée dans le chapitre VI évalue l'impact de cette transformation.

Il est important de remarquer que, de manière générale, la taille du début d'année n'est pas nécessairement appropriée pour évaluer la taille d'une entreprise qui connait par exemple une très forte croissance. Ce point est développé dans la sous-partie suivante (à propos du total actif). Mais dans le cadre de la variable VM, la dimension prospective des évaluations faites par le marché permet de négliger ce problème.

Enfin, une autre transformation peut être proposée. Au lieu de prendre la valeur de marché de l'année précédente, il est envisageable de prendre (VM-IB[63]). Autrement dit, utiliser la valeur de marché de fin d'année à laquelle on soustrait le résultat annuel. Une telle transformation est étudiée dans le cas de la variable AT. Mais dans le cas de la variable VM, il n'y a pas de sens à retrancher le résultat de la valeur de marché de fin d'année dans la mesure où la valeur de marché n'est pas une donnée comptable, mais une évaluation réalisée par les investisseurs. Ceux-ci utilisent notamment les données comptables pour réaliser leur estimation, mais pas uniquement. Retrancher le résultat de la valeur de marché ne permet donc pas d'obtenir une valeur de marché corrigée qui

[62] Degeorge, Patel et Zeckhauser, 1999 sont sans doute les seuls auteurs qui ne réduisent pas la variable résultat par action par un indicateur de taille.

[63] IB est l'abréviation retenue dans la thèse pour « résultat net » car elle correspond au nom de la variable « Income Before extraordinary items » dans la base de données utilisée (voir chapitre V).

serait celle établie par le marché si le résultat était nul. Le correctif qui est techniquement possible sur la valeur comptable qu'est l'actif total n'est pas approprié pour la valeur de marché.

b. L'actif total

Plusieurs auteurs (Bisson, Dumontier et Janin, 2004 ; Mard, 2004 ; Coppens et Peek, 2005) utilisent l'actif total (AT) comme dénominateur de « mise à l'échelle ». Le AT est une mesure comptable de la taille d'une entreprise, qui inclut le résultat. Il en découle donc que le numérateur et le dénominateur ne sont pas indépendants.

Pour isoler ces deux observations, plusieurs auteurs réduisent alors le résultat par l'actif total du début d'année, c'est-à-dire $AT_{(n-1)}$ dans la base (Kasznik, 1999 ; Payne et Robb, 2000 ; Bhattacharya, Daouk et Welker, 2003 ; Leuz, Nanda et Wysocki, 2003 ; Goncharov et Zimmermann, 2006 ; Petrovits, 2006 ; Roychowdhury, 2006 ; Shuto, 2008). Cependant, cette transformation pose deux problèmes méthodologiques :

(1) Puisqu'il utilise deux variables issues de deux exercices différents, ce dénominateur ne permet pas de calculer la distribution des résultats de la première année et réduit le nombre d'observations. De plus, la perte ne se limite pas à la première année, mais à la première année de chaque entreprise entrant dans la base. La description de la base menée dans le chapitre V évalue l'importance de ces données manquantes.

(2) Si une entreprise réalise de forts investissements durant l'année, l'indicateur de taille (actif total de l'année précédente) n'est plus pertinent pour mettre le résultat de l'année à l'échelle. Ce cas peut alors conduire à des valeurs extrêmes (résultat très important relativement au dénominateur). Une étude des cas extrêmes est donc menée dans le cadre de la description de la base de données (voir chapitre V) afin de mesurer l'impact de cette limite.

Pour limiter les effets indésirables de la réduction par l'actif total de l'année précédente, il est possible d'utiliser comme indicateur de taille la variable AT diminuée du montant du résultat de l'année en cours (IB). Aucun auteur n'utilise cette transformation. Pourtant elle a plusieurs avantages :

(1) Le dénominateur ainsi transformé est indépendant du dénominateur ; il n'inclut pas la variable étudiée comme c'est le cas si le résultat est divisé directement par la variable AT.

En cela, il répond à la préoccupation qui conduit à utiliser l'actif total de l'année précédente.

(2) Le principal avantage de ce ratio est d'utiliser des variables issues des états financiers de la même année. Cela diminue le nombre de données manquantes.

(3) Ce ratio réduit l'asymétrie de la distribution. En effet, il permet de tenir compte des variations d'investissement qui ont pu se produire durant l'année en cours et évite les cas extrêmes liés à l'utilisation d'un dénominateur fondé sur une structure financière obsolète. Ces cas peuvent sembler rares et l'analyse descriptive de la base menée dans le chapitre V permet de les identifier. Ils peuvent sembler négligeables dans la mesure où ils conduisent à des valeurs extrêmes du ratio (résultat important relativement au dénominateur) qui ne se situent donc pas à proximité du seuil résultat nul. L'impact de cette limite est donc négligeable dans le cadre de mesures non paramétriques qui ne s'intéressent qu'aux observations jouxtant le seuil. Il est cependant très important dans le cadre du travail mené dans la troisième partie de la thèse concernant la mesure paramétrique des irrégularités. Le Tableau 7 détaille cette analyse.

Tableau 7 : Comparaison des caractéristiques mathématiques des distributions des deux transformations de la variable AT utilisées comme indicateur de taille

		Ratio IB / $AT_{(-1)}$	Ratio IB / (AT-IB)
Résultat négatif	La perte est extrêmement élevée par rapport au montant des actifs (limite inférieure)	Le ratio tend vers -∞	Le ratio tend vers -1
	Le montant de la perte est (en valeur absolue) égal au montant des actifs	Le ratio est de l'ordre de -1	Le ratio est de l'ordre de -0,5
	Le montant de la perte est (en valeur absolue) égal à la moitié du montant des actifs	Le ratio est de l'ordre de -0,5	Le ratio est de l'ordre de -0,33
	Le montant de la perte est très faible (relativement au montant des actifs)	Le ratio tend vers -0	Le ratio tend vers -0
Résultat positif	Le montant du bénéfice est très faible (relativement au montant des actifs)	Le ratio tend vers +0	Le ratio tend vers +0
	Le montant du bénéfice est de l'ordre de la moitié du montant des actifs	Le ratio est de l'ordre de 0,5	Le ratio est de l'ordre de 1
	Le montant du bénéfice est égal au montant des actifs (limite supérieure)	Le ratio tend vers 1	Le ratio tend vers +∞
	Le montant du bénéfice est très supérieur au montant des actifs[64].	Le ratio peut aller (en théorie) jusqu'à +∞	Ce cas est impossible.

La comparaison des deux ratios permet de constater que l'immense majorité des observations (partie centrale de la distribution) est relativement similaire (centrée autour

[64] Ce cas correspond à une situation absurde qui n'est possible que parce que l'indicateur de taille n'est pas issu de la même année que le résultat. Cette situation peut se produire si par exemple des investissements importants ont été réalisés lors de l'année étudiée.

de 0). L'étude du seuil résultat nul n'est donc pas fondamentalement différente selon les deux transformations.

En revanche, des différences notables peuvent être observées en ce qui concerne le traitement des valeurs extrêmes. Le ratio utilisant (AT-IB) fait techniquement disparaître les valeurs extrêmes négatives qui sont présentes en utilisant $AT_{(-1)}$ et qui peuvent correspondre aux situations de grand bain comptable. À l'autre extrémité de la distribution, les valeurs extrêmes positives du ratio utilisant $AT_{(-1)}$ n'apparaissent pas en utilisant (AT-IB). Ces valeurs sont d'ailleurs absurdes : un résultat supérieur à la valeur des actifs n'a pas de sens ; elle est cependant possible parce que les deux variables sont issues de deux années différentes et parce que les investissements réalisés au cours de l'année étudiée ne figurent pas dans les comptes de l'année précédente.

Le chapitre V, qui présente la base de données, passe en revue les valeurs extrêmes des deux ratios afin d'évaluer dans quelle mesure ceux-ci peuvent véritablement avoir un impact sur l'étude des seuils. Dans le chapitre VI, l'étude empirique est menée avec les deux ratios afin de les comparer.

c. Le chiffre d'affaires

Glaum, Lichtblau et Lindemann (2004) justifient l'utilisation du chiffre d'affaires (variable CA) comme dénominateur de « mise à l'échelle » par le fait que cette variable est moins susceptible d'être affectée par les caractéristiques des systèmes comptables nationaux que les autres variables (capitaux propres ou total actifs). Dans le cadre d'une étude internationale, ce choix peut sembler judicieux. Mais s'il est possible que la variable CA soit moins sensible aux caractéristiques des systèmes comptables, il peut l'être davantage aux caractéristiques sectorielles.

Par ailleurs, comme dans le cas de la variable AT et de la variable VM, il est légitime de se poser la question de l'utilité de transformer la variable ; faut-il utiliser $CA_{(n-1)}$ ou (CA-IB) comme dénominateur ? Une transformation pourrait sembler à première vue d'autant plus utile que la part du résultat dans le chiffre d'affaires est généralement plus importante que dans l'actif total (2,1 % contre 0,6 % ; voir Tableau 8).

Tableau 8 : Comparaison de la part du résultat respectivement dans les variables AT, le CA et la VM

		1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	moyenne
Moyenne annuelle	IB/AT	0.44%	0.21%	0.32%	0.19%	0.42%	0.75%	1.06%	1.11%	1.41%	0.25%	-0.06%	0.76%	1.17%	**0.62%**
	IB/CA	1.57%	0.77%	1.14%	0.69%	1.54%	2.59%	3.50%	3.84%	4.37%	0.37%	-0.20%	2.69%	3.99%	**2.07%**
	IB/VM	3.11%	1.35%	2.72%	1.06%	3.12%	3.86%	4.09%	2.85%	4.75%	1.13%	-0.40%	4.32%	6.17%	2.93%
Médiane annuelle	IB/AT	1.57%	1.21%	1.92%	1.74%	2.06%	2.60%	3.02%	3.24%	2.80%	2.14%	1.74%	1.71%	2.67%	2.19%
	IB/CA	2.41%	2.23%	3.20%	2.51%	3.01%	3.55%	4.06%	4.09%	3.64%	2.45%	2.05%	2.19%	3.20%	2.97%
	IB/VM	2.60%	1.62%	3.30%	4.27%	4.23%	4.30%	3.87%	2.93%	3.40%	3.82%	4.68%	3.74%	4.41%	3.63%

Il n'apparaît pas nécessaire d'utiliser le chiffre d'affaires du début d'année. Le résultat étant partie intégrante du total actif, il est délicat de mettre le résultat à l'échelle par une variable qui intègre ce même résultat et c'est pourquoi on utilise l'actif total (comme la valeur de marché) du début d'année comme indicateur de taille de l'entreprise. Ce problème ne se pose pas de la même manière avec la variable chiffre d'affaires car l'activité d'une entreprise peut connaître des variations importantes d'une année sur l'autre, et $CA_{(n-1)}$ n'est pas nécessairement significatif de l'activité de l'année n. Le résultat de l'année n doit donc être comparé à l'activité qui a permis de générer ce résultat, donc du chiffre d'affaires de l'année n.

Par ailleurs, le résultat n'est pas nécessairement une composante du chiffre d'affaires. Une entreprise peut réaliser une perte (voire un bénéfice), même si son activité est nulle, du fait de l'existence de charges fixes, de sa structure financière ou de charges et produits exceptionnels par exemple. Diminuer le résultat du chiffre d'affaires ne fournit pas une information pertinente sur la taille de l'entreprise. Diminuer le résultat du chiffre d'affaires annuel comme cela a été fait pour l'actif total n'est pas pertinent.

2. Démarche de l'étude méthodologique

Les études qui s'intéressent aux irrégularités de distribution du résultat sont confrontées à un choix entre plusieurs variables ($VM_{(n-1)}$, $AT_{(n-1)}$, (AT-IB), CA) de mise à l'échelle. L'étude cherche à évaluer empiriquement l'impact de ce choix sur la mesure des irrégularités.

a. Démarche

Pour évaluer l'impact du choix du dénominateur sur la mesure de l'irrégularité, les valeurs des CA, AT et VM des entreprises présentes sur les intervalles d'observation jouxtant le

seuil du résultat nul sont comparées de manière à évaluer dans quelle mesure ces variables peuvent être responsables d'une distorsion significative. Puis les irrégularités sont mesurées et comparées année après année avec les trois dénominateurs sur la population de référence.

b. Résultats attendus de l'étude empirique sur l'impact du dénominateur

Si le dénominateur est source de distorsions, il est peu probable que les trois dénominateurs soient sources des mêmes distorsions. La mesure des corrélations entre les différentes variables de mise à l'échelle doit donc permettre d'évaluer l'impact des distorsions liées au dénominateur.

Si les trois variables sont fortement corrélées, cela signifie que les discontinuités, même si elles sont influencées par un effet dénominateur, peuvent être interprétées comme des discontinuités de la variable résultat. L'effet dénominateur peut être négligé.

Si les trois variables ne sont pas corrélées, cela signifie que l'effet de tous ou certains dénominateurs ne peut être négligé. Dans ce cas, l'étude doit proposer des critères de sélection de la meilleure variable de mise à l'échelle en fonction de la population étudiée, de la composition de la base utilisée et du seuil étudié.

CHAPITRE V – DÉMARCHE GÉNÉRALE DE L'ÉTUDE EMPIRIQUE ET ANALYSE DE LA BASE DE DONNÉES

Pour effectuer les tests empiriques, la base de données utilisée est constituée des entreprises françaises cotées dont les données financières sont présentes sur la base Global Vantage. Dans la section A, la base de données est décrite. Certaines corrections sont effectuées et sont décrites dans la section B. La section C décrit la base corrigée.

A. Analyse préliminaire de la base initiale

1. Constitution de la base

Dans la base Global Vantage (Compustat ; Standard & Poor's) constituée de sociétés cotées, sont retenues toutes les sociétés industrielles actives (code « $G » dans la base) et inactives (code « $D »), et toutes les sociétés financières actives (code « $F ») et inactives (code « $X ») françaises (code « Comp. Specifics country = FRA ») soit 1011 sociétés[65]. Les données accessibles de 1992 à 2004 couvrent près de 7746 observations.

a. Les entreprises inactives

Pour constituer la base, les entreprises actives et inactives de la base Global Vantage sont retenues afin d'éviter un biais de sélection en ne retenant que les entreprises survivantes. Ce faisant, en augmentant ainsi la population étudiée et le nombre d'observations, la proportion de données manquantes augmente inévitablement. En effet,

[65] L'extraction des données a été faite en mars 2006.

chaque année, une donnée peut être manquante soit parce que l'entreprise n'est pas encore présente dans la base, soit parce qu'elle en est sortie, soit parce que la base est incomplète. Autrement dit, le taux d'observations annuelles par rapport à la population totale ne doit pas être interprété uniquement comme un indicateur de données manquantes.

Un exemple permet d'illustrer ce propos. En 2005[66], le nombre de sociétés actives était de 643. En 2006, il était de 615. Entre 2005 et 2006, plus de 28 sociétés sont donc sorties de la base (sachant que d'autres y sont entrées) soit qu'elles aient quitté la cote, soit qu'elles aient été liquidées, soit qu'elles aient fusionné ou été rachetées. Le Tableau 9 montre que, si seules les 643 entreprises actives en 2005 avaient été retenues, le nombre d'observations depuis 1994 aurait été de 5 413. Chaque année, il ne peut en aucun cas être supérieur à 643, et il diminue naturellement en remontant le temps.

Tableau 9 : Base constituée sur les seules sociétés actives en 2005 (biais du survivant)

Année	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Nombre d'observations parmi les 643 entreprises actives en 2005 (Na)	206	226	336	425	499	591	625	626	631	618	630	**5 413**
Na/643	32%	35%	52%	66%	78%	92%	97%	97%	98%	96%	98%	

En comparaison, en élargissant la base aux 1011 sociétés actives et inactives, le nombre d'observations (sur la même période de 1994 à 2004) s'élève à 7 041, soit une augmentation de près de 30 %[67].

Tableau 10 : Base constituée sur les sociétés actives et inactives de 1992 à 2005

Année	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Population (P)	1011	1011	1011	1011	1011	1011	1011	1011	1011	1011	1011	11 121
Nombre d'observations (N) de la variable IB dans la base non corrigée	377	423	617	693	741	782	779	677	696	655	601	**7 041**
N/P	37%	42%	61%	69%	73%	**77%**	**77%**	67%	69%	65%	59%	63%
Écart avec les sociétés actives (N-Na)	171	197	281	268	242	191	154	51	65	37	29	**1 628**

En conclusion, le taux d'observation par an, qui ne dépasse jamais les 77 % de la base, est essentiellement dû aux variations de la constitution de la base (entrées et sorties de cote). Les véritables données manquantes sont relativement peu nombreuses chaque année.

[66] Mars 2005 correspond à la date d'extraction de la base utilisée pour certains travaux exploratoires.

[67] (7041 - 5413)/5413 = 1628/5431 = 30,08 %.

b. Les entreprises de banque et assurance

Les études statistiques en comptabilité qui utilisent des bases de données ont pour habitude de supprimer les entreprises de banque et assurance. Cela se justifie du fait que les règles comptables qui leur sont appliquées sont différentes, et que l'interprétation, et donc la comparaison, des chiffres (aussi bien de structure du bilan que de performance) sont délicates.

Dans le cas d'une étude sur les irrégularités de distribution, la pertinence de cet usage peut être questionnée. En effet, ce n'est ni la performance des entreprises, ni leur structure financière dans l'absolu qui est comparée, mais leur comportement (collectif) face à un seuil de résultat publié. Or ce comportement vis-à-vis du seuil peut se manifester aussi bien chez les entreprises industrielles que les entreprises financières, indépendamment des règles comptables qu'elles doivent appliquer. Si le secteur peut servir de variable de contrôle, il ne semble pas pour autant nécessaire de supprimer les entreprises financières de la base de données[68].

c. Choix des Variables

Mard (2004) étudie 27 rapports annuels de sociétés cotées françaises en 1998. Il répertorie les informations sur les résultats contenues dans ces rapports. Il observe que le résultat net est le solde le plus souvent mis en évidence dans les rapports annuels (dans 93 % des cas). Viennent ensuite le résultat d'exploitation (présent dans 70 % des rapports), le résultat net par action (48 %) et le résultat courant (30 %). En conclusion, la variable dont la distribution est retenue comme objet de l'étude est le résultat net, car c'est la variable de performance sur laquelle les entreprises communiquent le plus.

Dans la base Global Vantage, le résultat net correspond à la variable « IB[69] ». Cependant, dans un souci d'harmonisation internationale, ce résultat peut être diminué de certains résultats extraordinaires qui ne correspondent pas aux résultats exceptionnels en comptabilité française. Ces résultats extraordinaires sont en général nuls. Lorsqu'ils apparaissent, ils sont presque toujours négligeables, à quelques rares exceptions près comme par exemple le résultat net de Gaz De France en 2001 qui s'élève à 3 612 000 000 € en raison essentiellement du rachat du réseau de transport par l'État. Ces cas sont cependant extrêmement rares et la variable IB est donc assimilable au

[68] 104 entreprises de la base portent le code secteur financier (« sector code » 5000 dans la base), soit 10,27% de la population.

[69] En anglais, IB est l'abréviation de « Income Before extraordinary items ».

résultat net habituellement publié par les entreprises françaises. Un retraitement systématique ne se justifie pas dans le cadre d'une étude qui ne s'intéresse essentiellement qu'aux très faibles bénéfices et aux très faibles pertes.

d. Mise à l'échelle[70]

Le résultat (en valeur) d'une entreprise n'a pas le même sens en fonction de la taille de l'entreprise (Burgstahler et Dichev, 1997 ; Degeorge, Patel et Zeckhauser, 1999), notamment lorsque l'on s'intéresse à la notion de « petite » perte ou de « petit » gain. La distribution de la variable résultat net est donc mise à l'échelle par une variable de taille. Plusieurs variables de taille sont utilisées dans l'étude : Chiffre d'Affaires (CA), Total Actif (AT), ou valeur de marché (VM).

Dans la base Global, les variables IB (pour résultat net), AT[71] (pour actif total), REVT[72] (pour chiffre d'affaires), et MKT[73] (pour valeur de marché) ont donc été extraites. Par ailleurs, les variables descriptives ECNSC[74] (pour secteur économique), et GVKEY (pour « Global Vantage Company Key ») ont également été extraites.

e. Données internationales

Seules les entreprises françaises sont considérées dans cette partie de l'étude. Dans la partie 3, l'étude est étendue à des entreprises allemandes et anglaises. C'est pour cette raison que la base de données internationale Global Vantage est retenue.

Les mêmes informations ont donc été extraites pour les entreprises françaises (1011 entreprises), britanniques (2094 entreprises) et allemandes (981 entreprises)[75].

f. Données trimestrielles

Une tentative a été menée pour extraire les résultats trimestriels, mais il est rapidement apparu que ces informations sont inexploitables. En effet, dans la quasi-totalité des cas, le

[70] En anglais : « scaling ». Nous aurions pu utiliser l'expression « réduction », mais ce terme sera utilisé dans la partie 3 lorsque les données sont centrées et réduites. Pour ne pas créer de confusion, nous parlons donc de « mise à l'échelle ».

[71] En anglais : « Assets Total ».

[72] En anglais : « Revenue Total ».

[73] En anglais : « Market Value – total ».

[74] En anglais : « Economic sector code ».

[75] Dans la base, la variable de nationalité est CINC (pour « Country of Incorporation ») = GBR et DEU.

résultat trimestriel inscrit est le résultat annuel divisé par 4. L'information n'est donc pas pertinente.

2. Synthèse de la base initiale

Le Tableau 11 ci-dessous présente la base de données avant ajustements.

Tableau 11 : Synthèse des caractéristiques de la base de données (entreprises françaises) avant ajustement.

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	total période
Nombre d'entreprises dans la base	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	1 011	14 154 observations potentielles

Résultat Net

Nombre d'observations	291	295	377	423	617	693	741	782	779	677	696	655	601	42	7 669
% de données manquantes + sociétés inactives	70.8%[76]	62.7%	58.1%	38.9%	31.4%	26.7%	22.6%	25.8%	33.0%	22.9%	31.1%	35.2%	40.5%	95.8%	45.8%

Total Actif

Nombre d'observations	292	296	376	422	617	699	752	801	781	733	697	657	601	44	7 768 observations réelles
Moyenne	9 421.2	10 210.2	8 904.3	8 657.5	6 445.0	5 910.0	5 546.4	5 859.2	6 965.5	7 725.0	7 964.4	8 646.5	9 161.8	29 556.4	AT moyen 7 581[77]
Écart type	32 964	35 753	31 775	32 443	30 229	32 044	31 971	39 443	43 397	48 546	48 262	55 858	61 175	189 524	
Max	282 027	300 691	264 508	290 024	378 401	383 362	393 777	698 625	694 037	825 296	710 319	785 985	905 938	1 258 079	
Min	3.96	8.29	8.74	7.81	3.10	1.79	0.64	1.09	0.00	0.17	0.04	0.06	2.26	6.45	

[76] 70,8 % = (1011 – 291) / 1011.

[77] L'actif total moyen calculé de 1992 à 2005 est 7 581 millions d'euros (moyenne de la ligne pondérée par le nombre d'entreprises annuel). Mais comme le nombre de données manquantes est très important en 2005, le même calcul a été effectué sur la seule période 1992 à 2004. L'actif total moyen est alors de 7 455 millions d'euros. La différence est faible (moins de 2 %).

B. Correction de la base

Un certain nombre de tests ont été menés dans le but de corriger des informations manifestement erronées présentes dans la base. Ces corrections concernent essentiellement la variable résultat net des entreprises françaises qui est au cœur des préoccupations de l'étude. Ces corrections sont détaillées en Annexe 1 : Correction de la base.

1. Résultat strictement égal à zéro

Dans un premier temps, les résultats strictement égaux à zéro ont été examinés. Neuf données ont été supprimées, une a été corrigée, huit données manquantes identifiées ont été corrigées, et une société a été supprimée de la base.

2. Actif total strictement égal a zéro

Dans un deuxième temps, le même examen a été réalisé sur la variable actif total. Une seule donnée a été corrigée.

3. Données récentes

Dans un troisième temps, les données de l'année 2005 ont été examinées. Comme ces données sont trop peu nombreuses, l'année 2005 a été supprimée de la base. L'horizon d'analyse est donc de 13 ans (1992 à 2004).

4. Données manquantes

Dans un quatrième temps, un travail a été mené afin de compléter les données manquantes de la base dans la mesure du possible. Ce travail s'est avéré chronophage, mais il permet de récupérer près de 200 observations (soit près de 2,5 % de la population). Les données manquantes sont identifiées de deux manières : (1) en identifiant les « trous » dans les séries de résultats, et (2) en vérifiant la présence simultanée de la variable résultat avec une variable de mise à l'échelle.

En ce qui concerne les « trous », les cas étranges sont identifiés par un test sous Excel consistant à identifier les absences de résultat une année lorsque le résultat est présent dans la base l'année d'avant, et l'année d'après. Les données sont ensuite complétées, société par société, en recherchant les informations directement dans les rapports financiers (téléchargés sur Internet) ou dans la base Thomson Worldscope. Ce travail de collecte est long, fastidieux, et pas nécessairement très productif. Il permet cependant de compléter 73 données, et une nouvelle société est supprimée de la base (qui comporte désormais 1009 sociétés).

En ce qui concerne la présence simultanée de la variable résultat avec une variable de mise à l'échelle, il apparaît que pour 53 sociétés, les données manquantes correspondent à des années où le résultat de l'entreprise est connu, mais pas son total actif, ou inversement. Au total, 133 données sont complétées en retrouvant l'information dans les rapports annuels disponibles sur Internet ou dans la base Thomson.

5. Valeurs extrêmes

Dans un cinquième temps, dans la mesure où l'étude s'intéresse essentiellement à la distribution de la variable résultat divisé par une variable de taille, une attention particulière est portée à certaines valeurs très extrêmes. Les valeurs extrêmes des ratios $IB/AT_{(n-1)}$ et IB/(AT-IB) sont examinées successivement.

Dans le premier cas (ratio $IB/AT_{(n-1)}$), sont appelées valeurs « très extrêmes » les valeurs supérieures à 1 ou inférieures à -1. En effet, cela signifie que l'entreprise réalise des bénéfices (pertes) supérieur(e)s au montant de l'actif. Ces valeurs extrêmes peuvent avoir deux causes :

1. Une erreur dans la base, qu'il est utile de corriger.

2. Une situation financière extrême justifiée par une conjoncture économique spécifique. Dans ce dernier cas, aucune correction ne s'impose. De telles valeurs extrêmes[78] ont seulement été répertoriées.

Dans le second cas (ratio IB/(AT-IB)), il apparaît que l'utilisation de ce ratio permet de limiter (ou de « tasser ») les valeurs extrêmes négatives, sans modifier de manière significative la partie centrale de la distribution, sur laquelle figure le seuil d'étude. Le seul

[78] Les valeurs extrêmes ont des conséquences importantes sur les mesures paramétriques des irrégularités. Elles font l'objet d'un second examen dans la partie 3.

cas de figure atypique serait constitué par d'éventuelles valeurs extrêmes positives, mais on sait par ailleurs qu'elles sont rares, et jamais aussi extrêmes que les valeurs négatives.

Ainsi, sur l'échantillon, il n'apparaît que 6 valeurs extrêmes positives (supérieures à 1). À une exception près, les valeurs ne sont jamais supérieures à 2. Dans tous les cas, elles apparaissent comme des évènements ponctuels que rien ne permet d'interpréter comme étant des valeurs aberrantes.

Ces cas extrêmes font l'objet d'un examen plus approfondi dans la partie 3. Dans le cadre de la partie 2, ils n'ont pas d'impact sur les mesures non paramétriques utilisées.

C. Analyse descriptive de la base corrigée

1. Effectifs de l'échantillon

Mis à part dans le cadre de la section C du chapitre VI qui étudie l'influence des différents dénominateurs, l'étude s'intéresse essentiellement à la variable résultat (IB) mise à l'échelle par l'actif total (plus précisément par AT-IB). Après l'élimination des deux sociétés et les corrections apportées aux données décrites dans la section précédente et détaillées en Annexe 1 : Correction de la base, la taille de la population étudiée apparaît ainsi :

Tableau 12 : Composition de la base après corrections ; variable IB / (AT-IB)

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Population	1009	1009	1009	1009	1009	1009	1009	1009	1009	1009	1009	1009	1009
Nombre d'observations	291	296	376	422	616	702	753	798	791	734	698	659	610

2. Évolution de la moyenne, la médiane et l'écart type

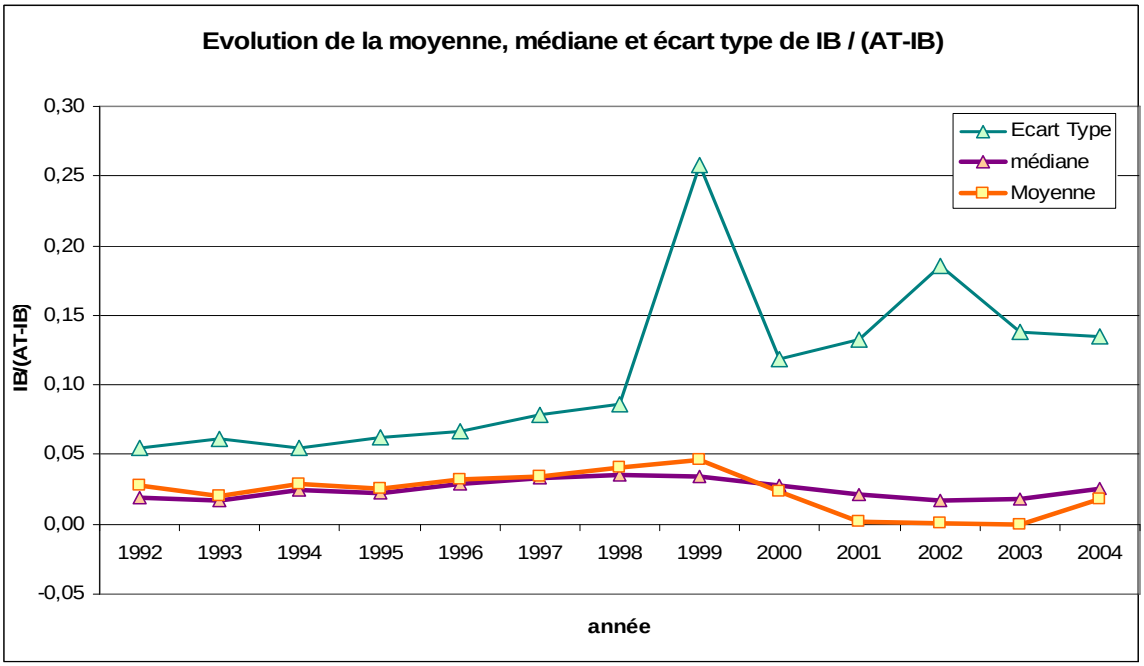
L'analyse descriptive de la base se concentre sur l'examen de la moyenne, de l'écart type, de la médiane, et des valeurs extrêmes tels qu'ils apparaissent dans le Tableau 13.

Tableau 13 : Caractéristiques de la base après corrections ; variable IB / (AT-IB)

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Moyenne	0.02796	0.01991	0.02895	0.02534	0.03212	0.03407	0.04057	0.04630	0.02334	0.00135	0.00123	-0.00020	0.01778
Médiane	0.01967	0.01733	0.02488	0.02270	0.02848	0.03328	0.03496	0.03459	0.02807	0.02143	0.01717	0.01845	0.02592
Intervalle Inter Quartile	0.02919	0.03462	0.02860	0.03166	0.03125	0.02727	0.03634	0.03725	0.03287	0.03063	0.03113	0.02974	0.03052
Écart type	0.05447	0.06160	0.05445	0.06183	0.06652	0.07909	0.08651	0.25799	0.11852	0.13286	0.18550	0.13754	0.13500
Max	0.21356	0.23166	0.29045	0.24932	0.46349	0.58659	0.65051	6.81741	0.80137	0.58370	1.98373	1.80595	1.80919
Min	-0.22244	-0.31440	-0.31026	-0.42143	-0.49237	-0.70967	-0.58155	-0.46536	-0.80061	-0.79830	-0.98091	-0.59434	-0.73009

Pour plus de lisibilité, ces valeurs sont représentées graphiquement dans les Figure 37 et Figure 38.

Figure 37 : Évolution de moyenne, médiane, et écart-type de la variable IB / (AT-IB)



a. La moyenne

Une certaine stabilité de la performance économique moyenne des entreprises entre 1992 et 1995 est constatée, suivie d'une amélioration entre 1996 et 1999, puis par une dégradation dès 2000 qui s'accentue très fortement entre 2001 et 2003.

Hormis en 2003, la moyenne est toujours positive. Cela signifie que sur les distributions annuelles des résultats qui font l'objet de notre étude, le sommet de la courbe (qui est autour de la moyenne) est toujours situé à droite du résultat nul. Autrement dit, le seuil se situe presque toujours sur la partie gauche de la distribution, à l'exception de 2001 et de 2003 où il se situe très proche du sommet de la distribution (la moyenne est quasiment

égale à zéro). Cette observation mérite d'être relevée car elle justifie le fait que l'étude se concentre sur la partie gauche de la courbe de distribution. La partie 3 de la thèse développe ce point.

b. L'écart type

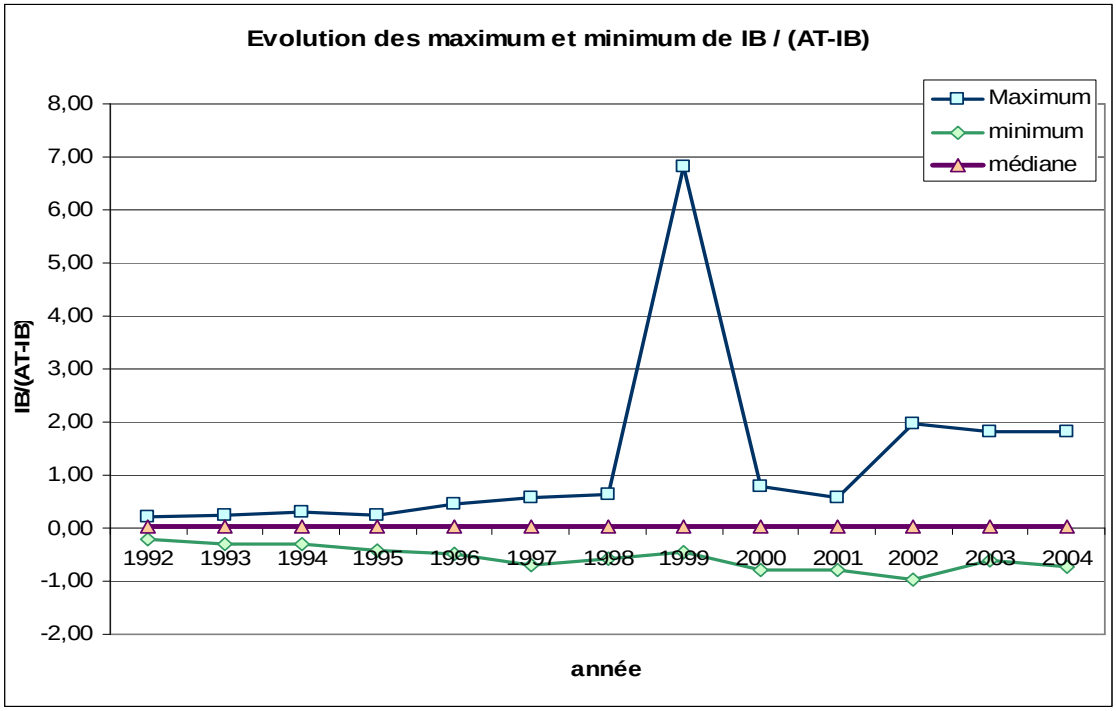
Il apparaît une augmentation assez régulière de l'écart type entre 1992 et 2002. Il apparaît un « pic » en 1999 dû à une seule variable positive extrême (Systran). Après avoir atteint un maximum en 2002, il se stabilise à un niveau plus modéré. La période 2000 à 2003 correspond à une phase économique d'incertitudes et de turbulences qui explique la faiblesse de la performance moyenne et la progression de l'écart type.

c. La médiane

La médiane évolue sensiblement de la même manière que la moyenne. Mais elle présente un caractère plus « lisse » notamment durant les années pour lesquelles l'écart type est élevé. En effet, par construction, la médiane est beaucoup moins sensible aux valeurs extrêmes qui altèrent écart type et moyenne.

d. Maximum et minimum

Figure 38 : Évolution de max et min de la variable IB / (AT-IB)



La Figure 38 permet de constater que ce sont essentiellement les (rares) valeurs extrêmes qui expliquent l'envolée de l'écart type en 1999 et 2002.

3. Composition de la base

Tableau 14 : Composition sectorielle de l'échantillon

Code	Secteur Économique		
1000	Matières premières (Materials)	57	5.6%
2000	Consommation en équipements et services (Consumers Discretionary[79])	137	13.6%
3000	Consommation de produits courants (Consumers staples)[80]	64	6.3%
3500	Santé (Health care)	31	3.1%
4000	Énergie (Energy)	10	1.0%
5000	Services financiers (Financials)	104	10.3%
6000	Industrie (Industry)	158	15.6%
8000	Technologies de l'information (Information technology)	133	13.2%
9000	Services collectifs (Utilities)	8	0.8%
	Non renseigné	309	30.6%
		1 011	

Tous les secteurs d'activité sont représentés dans la base d'entreprises françaises cotées étudiée.

[79] Plus précisément, ce secteur inclut l'automobile, les biens durables d'équipement ménager, le textile, l'équipement de loisir, la restauration et l'hôtelerie, les médias, et les services.

[80] Plus précisément, ce secteur inclut toutes les industries qui produisent ou vendent alimentation, médicaments, tabac, et produits ménagers.

CHAPITRE VI – MISE EN ŒUVRE DE L'ÉTUDE MÉTHODOLOGIQUE

Par souci de symétrie, ce chapitre est construit selon le même plan que le chapitre IV. Les sections A, B et C répondent aux trois questions méthodologiques soulevées dans les sections A, B et C du chapitre IV : Quel est l'impact des méthodes de mesure, de la largeur des intervalles, et de la variable d'échelle sur l'identification et la mesure des seuils comptables.

A. Impact de la méthode de mesure de l'irrégularité

Les méthodes de mesure peuvent être divisées en deux familles. La première regroupe les mesures qui se fondent sur une estimation de l'effectif attendu pour en déduire un effectif irrégulier. La seconde regroupe les indicateurs d'asymétrie qui se contentent de comparer les effectifs observés de deux intervalles. L'étude examine successivement ces deux familles de mesures.

1. Étude des mesures fondées sur une estimation

L'étude cherche à mettre en évidence l'impact de la méthode adoptée pour estimer l'irrégularité. Les différentes mesures et les postulats implicites sur lesquels elles reposent sont décrits dans la section A du chapitre IV. Le Tableau 15 récapitule ces mesures en les classant par méthodologie mathématique.

Tableau 15 : Mesures non paramétriques des irrégularités. Classement par méthodologie mathématique.

Famille de mesure	Méthode	Commentaires		Avantages et limites
Symétries	Symétrie par rapport à la médiane		1	Voir Tableau 5 : Mesures non paramétriques des irrégularités. Postulats implicites et synthèse des limites et avantages
	Symétrie par rapport à la moyenne		2	
Moyennes	Moyenne arithmétique		3	
	Moyenne pondérée	Non étudié empiriquement dans la thèse		
	Moyenne pondérée corrigée			
Interpolations	Interpolation linéaire	Retrait du seul intervalle étudié	4	
		Estimation en retirant les deux intervalles irréguliers	5	
	Interpolation exponentielle	Retrait du seul intervalle étudié	6	
		Estimation en retirant les deux intervalles irréguliers	7	
	Interpolation logarithmique	Retrait du seul intervalle étudié	8	
		Estimation en retirant les deux intervalles irréguliers	9	
Extrapolations	Extrapolation linéaire	Intéressant pour estimer des effectifs situés proches du sommet de la courbe de distribution	10	
	Extrapolation logarithmique		11	
	Extrapolation exponentielle	Non étudié empiriquement dans la thèse		

En définitive, pour comparer les différentes mesures, chaque irrégularité (droite et gauche) est mesurée selon 11 méthodes différentes, sur 13 années, soit 268[81] mesures.

a. Intervalle d'étude

a.1. Choix de la largeur de l'intervalle

Les mesures non paramétriques de l'irrégularité imposent de déterminer une valeur de l'intervalle d'étude. Chaque année, la population étudiée évolue. Le nombre d'observations, l'écart type et l'intervalle interquartile de la distribution de résultat sont différents chaque année. Il est alors possible (1) de mesurer pour chaque année une largeur d'intervalle propre en utilisant une formule théorique (Silverman, 1986 ; Scott, 1992)[82], (2) ou de fixer une largeur d'intervalle arbitraire commune à toutes les années.

Dans le second cas, fixer une largeur d'intervalle unique pour toutes les distributions impose un second choix : (2a) la fixer arbitrairement, ou (2b) la mesurer par une formule

[81] 2 irrégularités x 11 méthodes x 13 années = 286 mesures.

[82] $2(IQR)n^{-1/3}$

théorique (Silverman, 1986 ; Scott, 1992) appliquée à la distribution agrégée de toutes les années confondues.

Tableau 16 : Caractéristiques de la distribution agrégée des 13 années, et largeur théorique de l'intervalle d'étude

IB/(AT-IB) de 1992 à 2004	
N	7 746
Min	-0.9809082
Max	6.817411
Moyenne	0.0229601
Médiane	0.0263726
IQR	0.055818
Largeur théorique de l'intervalle ($2iqrN^{-1/3}$)	**0.005642156**

Il apparaît que la largeur d'intervalle théorique mesurée sur l'agrégation des 13 distributions est très proche de 0,005. La largeur de 0,005 est également la plus généralement retenue par les auteurs en comptabilité (Burgstahler et Dichev, 1997 ; Burgstahler et Eames, 2003 ; Dechow, Richardson et Tuna, 2003 ; Mard, 2004 ; Brown et Caylor, 2005 ; Coppens et Peek, 2005 ; Durtschi et Easton, 2005 ; Beaver, McNichols et Nelson, 2007). Ceci étant dit, cette largeur théorique a été mesurée sur la distribution agrégée, or chaque distribution annuelle est considérée comme étant indépendante. Une largeur théorique pour chaque année a donc été calculée. Les résultats sont présentés dans le Tableau 16.

Tableau 17 : Caractéristiques des 13 distributions, et largeur théorique de l'intervalle d'étude

	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
N	291	296	376	422	616	702	753	798	791	734	698	659	610
min	-0.222	-0.314	-0.310	-0.421	-0.492	-0.710	-0.582	-0.465	-0.801	-0.798	-0.981	-0.594	-0.730
max	0.214	0.232	0.290	0.249	0.463	0.587	0.651	6.817	0.801	0.584	1.984	1.806	1.809
p50	0.020	0.017	0.025	0.023	0.028	0.033	0.035	0.035	0.028	0.021	0.017	0.018	0.026
mean	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0463	0.0233	0.0014	0.0012	-0.0002	0.0178
iqr	0.048	0.052	0.050	0.051	0.054	0.054	0.063	0.063	0.056	0.060	0.064	0.061	0.054
Largeur théorique de l'intervalle ($2iqrN^{-1/3}$)	**0.0144**	**0.0155**	**0.0139**	**0.0137**	**0.0128**	**0.0122**	**0.0139**	**0.0136**	**0.0121**	**0.0133**	**0.0144**	**0.0141**	**0.0128**

La taille annuelle (n) de l'échantillon étant nécessairement plus faible que la taille des échantillons agrégés, la largeur annuelle théorique est plus élevée. Elle oscille entre 0,0121 et 0,0155 (les largeurs théoriques sont toutes très proches et leur moyenne est de 0,01359231), soit de l'ordre de 2 à 3 fois la largeur la plus utilisée dans la littérature (0,005).

En définitive, comme l'étude de l'impact de la largeur des intervalles sur les mesures d'effet de seuil fait l'objet d'un développement ultérieur[83], il a semblé suffisant dans le cadre de ce premier axe de recherche d'établir une largeur d'intervalle unique et arbitraire, commune à toutes les distributions. La largeur de 0,005 est donc retenue pour trois raisons. (1) Elle fait référence à la littérature antérieure. (2) Ce faisant, une « petite perte » ou un « petit bénéfice » est défini comme étant de l'ordre de la moitié de la largeur annuelle théorique. Il est important en effet que la largeur, même décidée arbitrairement, ne soit pas supérieure à la largeur théorique. (3) Le choix d'une valeur unique a également un autre avantage : il est important de prendre un chiffre rond pour que les limites entre intervalles soient plus simples à interpréter, et notamment que le seuil du résultat nul soit situé à la frontière entre deux intervalles.

a.2. Effectif des intervalles étudiés

Tableau 18 : Tableau récapitulatif des effectifs autour du seuil résultat nul

	[-0,025 ;-0,02[	[-0,02 ;-0,015[	[-0,015 ;-0,01[	[-0,01 ;-0,005[	[-0,005 ;0[	[0 ; 0,005[	[0,005 ; 0,01[	[0,01 ; 0,015[	[0,015 ; 0,02[	[0,02 ; 0,025[	TOTAL	Population	% de la population
1992	1	7	8	5	**14**	**32**	18	17	19	18	139	291	47.8%
1993	7	6	8	5	**5**	**38**	18	13	22	11	133	296	44.9%
1994	7	6	4	7	**10**	**41**	21	24	19	16	155	376	41.2%
1995	5	6	7	5	**13**	**41**	24	25	23	29	178	422	42.2%
1996	4	7	5	18	**15**	**44**	35	36	33	33	230	616	37.3%
1997	2	6	7	13	**21**	**47**	36	26	29	36	223	702	31.8%
1998	4	6	2	7	**9**	**46**	50	36	35	34	229	753	30.4%
1999	10	8	8	6	**8**	**35**	39	37	37	33	221	798	27.7%
2000	9	8	9	6	**9**	**35**	59	37	30	49	251	791	31.7%
2001	8	6	7	9	**13**	**42**	45	26	43	38	237	734	32.3%
2002	4	12	12	13	**11**	**42**	54	30	31	31	240	698	34.4%
2003	9	3	11	11	**9**	**34**	43	35	33	28	216	659	32.8%
2004	2	7	7	9	**3**	**26**	29	35	33	36	187	610	30.7%
											2639	**7746**	**34.1%**

Pour réaliser les calculs par interpolation, les effectifs doivent être relevés sur les 4 intervalles à droite et à gauche des intervalles étudiés. En ajoutant les deux intervalles d'étude (adjacents au seuil résultat nul) les effectifs sont relevés sur 10 intervalles autour du seuil. Au total, près d'un tiers de la population totale contribue aux estimations.

[83] Voir la section VI-B.

b. Détail des calculs

Cette sous-partie réunit quelques commentaires sur le processus de mesure des irrégularités

Mesures par moyenne

Pour mesurer l'effectif théorique [-0,005 ;0[il faut faire la moyenne de l'effectif [-0,01 ; -0,005[et [0 ; 0,005[. Pour mesurer l'effectif théorique [0 ; 0,005[il faut faire la moyenne de l'effectif [-0,005 ; 0[et [0,005 ; 0,01[. Ces calculs ne présentent pas de difficulté particulière.

Mesures par symétrie

Pour la mesure par symétrie, il faut relever les effectifs des intervalles autour de 2 fois la médiane et 2 fois la moyenne. Ces effectifs (Tableau 18) viennent s'ajouter aux précédents.

Figure 39 : Observations nécessaires à la mesure par symétrie

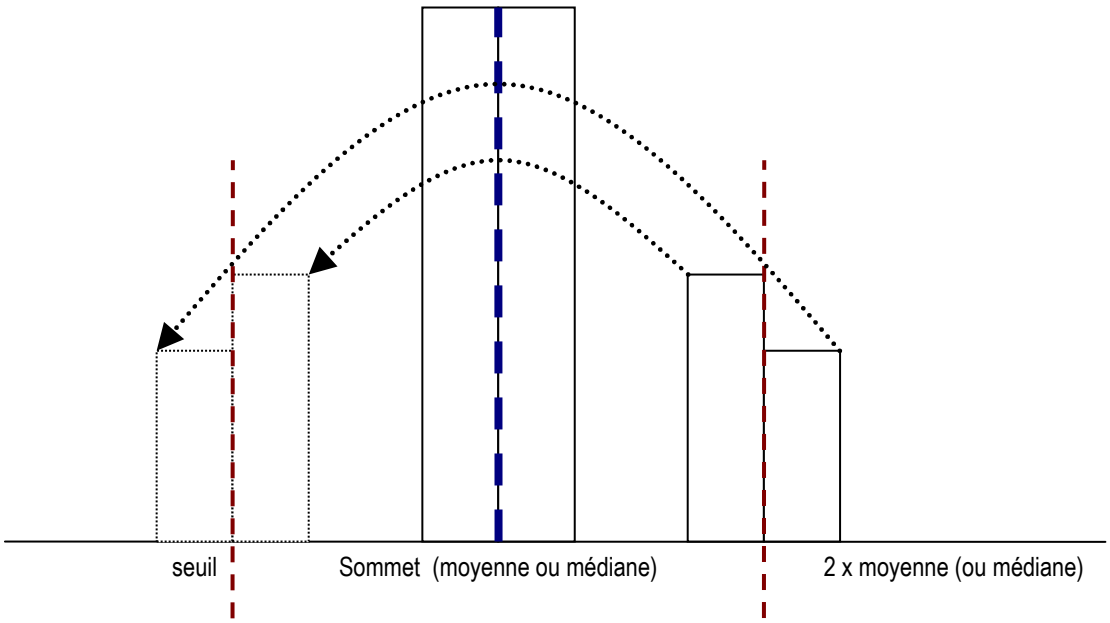


Tableau 19 : Tableau récapitulatif des effectifs autour de médiane x 2 et Moyenne x 2

m = médiane		M = Moyenne	
[2m-0,005 ; 2m[	**[2m; 2m+0,005[**	**[2M-0,005 ; 2M[**	**[2M; 2M+0,005[**
11	6	7	3
9	6	7	12
18	14	11	11
16	13	10	9
23	24	19	12
18	12	16	13
21	15	17	8
26	21	15	9
26	30	25	26
34	22	28	43
28	24	30	40
31	19	8	33
19	23	27	30

L'effectif de [-0,005 ;0[est estimé par l'effectif de [2M ; 2M+0,005[. L'effectif de [0 ; 0,005 [est estimé par l'effectif de [2M - 0,005 ; 2M [[84].

Mesures par interpolation

Les mesures par interpolation posent deux problèmes. Le premier est de décider s'il faut retirer uniquement l'intervalle à estimer de l'interpolation, ou les deux intervalles présumés irréguliers. En effet, comme il y a une double irrégularité, il semble difficilement justifiable d'estimer un intervalle irrégulier en incluant dans la distribution de référence l'intervalle adjacent lui-même présumé irrégulier. Pourtant, dans la littérature comptable, une telle correction n'est jamais réalisée.

Le second est de déterminer le nombre des intervalles à retenir pour effectuer l'interpolation. Dechow, Richardson et Tuna (2003) effectuent l'interpolation sur 4 intervalles avant et 4 intervalles après l'intervalle à estimer, soit une largeur totale de 9 intervalles, (une largeur de 0,045)[85]. L'étude s'inspire de cette méthodologie, mais en élargissant l'intervalle à 10 intervalles pour diminuer le risque d'estimer l'effectif d'un intervalle présumé irrégulière par l'effectif de l'intervalle adjacent lui-même présumé irrégulier. Autrement dit, l'interpolation se fait sur un intervalle global de 0,05.

[84] Idem avec la médiane « m ».

[85] Il est important de remarquer que c'est moins le nombre d'intervalles qui est important que la largeur totale correspondante dans la distribution, puisque chaque étude ne retient pas nécessairement la même largeur d'intervalle.

Quatre séries d'estimations ont été effectuées:

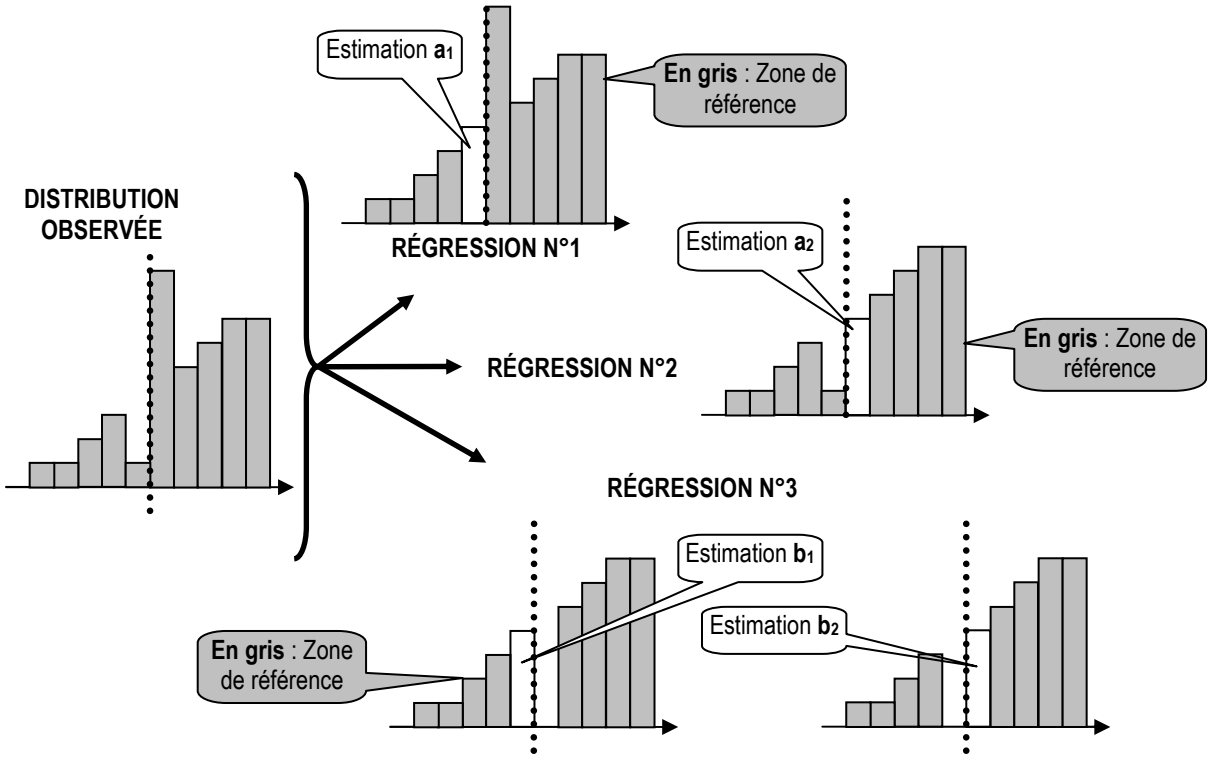
- Estimation de l'effectif des entreprises faiblement déficitaires sans retirer l'effectif des entreprises faiblement bénéficiaires des intervalles de référence (mesure a_1).

- Estimation de l'effectif des entreprises faiblement bénéficiaires sans retirer l'effectif des entreprises faiblement déficitaires des intervalles de référence (mesure a_2).

- Estimation des effectifs des entreprises faiblement bénéficiaires en retirant les deux intervalles entourant le seuil des intervalles de référence (mesures b_1).

- Estimation des effectifs des entreprises faiblement déficitaires en retirant les deux intervalles entourant le seuil des intervalles de référence (mesures b_2).

Figure 40 : Illustration des quatre séries de mesures par interpolation réalisées

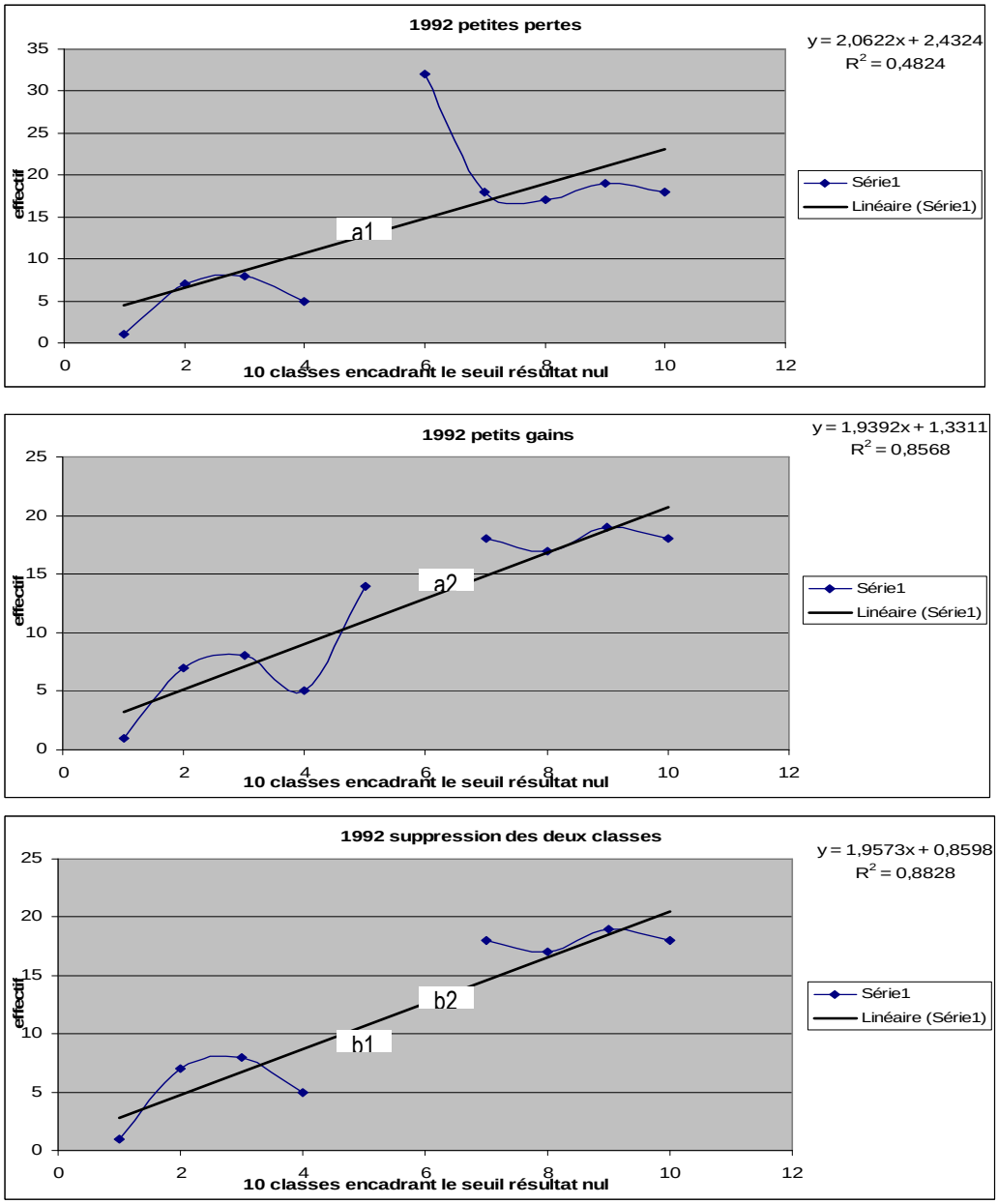


Puis pour chacune de ces séries, ont été effectuées trois interpolations différentes : linéaire, exponentielle, et logarithmique. Autrement dit, chaque année, 3 régressions sont effectuées, qui permettent d'effectuer 4 mesures chacune. Donc sur les 13 années, ce sont 9 x 13 = 117 régressions qui sont effectuées, permettant de faire 12 x 13 = 156 mesures au total.

Pour illustrer cette démarche et à titre d'exemple, la représentation graphique des interpolations est reproduite pour la seule année 1992. La Figure 41 correspond à

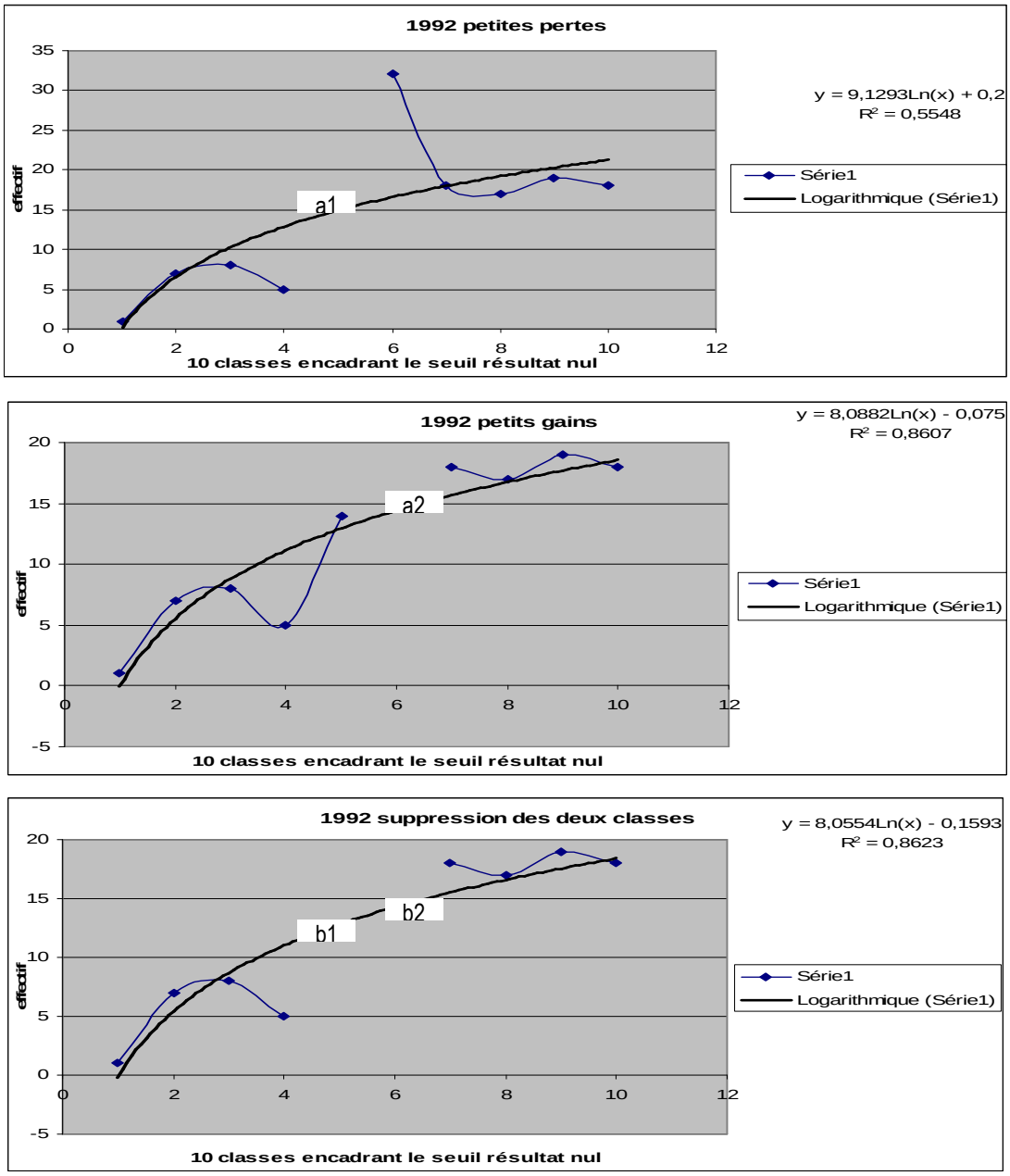
l'estimation par interpolation linéaire de l'effectif des entreprises faiblement déficitaires (premier graphique) et des entreprises faiblement bénéficiaires (deuxième graphique) en n'excluant de la régression que le seul intervalle estimé (a1 et a2 dans la Figure 40). Le troisième graphique de la Figure 41 représente la régression linéaire effectuée en excluant les deux intervalles irréguliers. Cette troisième interpolation permet d'estimer à nouveau aussi bien l'effectif des entreprises faiblement déficitaires que celui des entreprises faiblement bénéficiaires (b1 et b2 dans la Figure 40).

Figure 41 : Illustration graphique des interpolations linéaires menées sur la seule année 1992



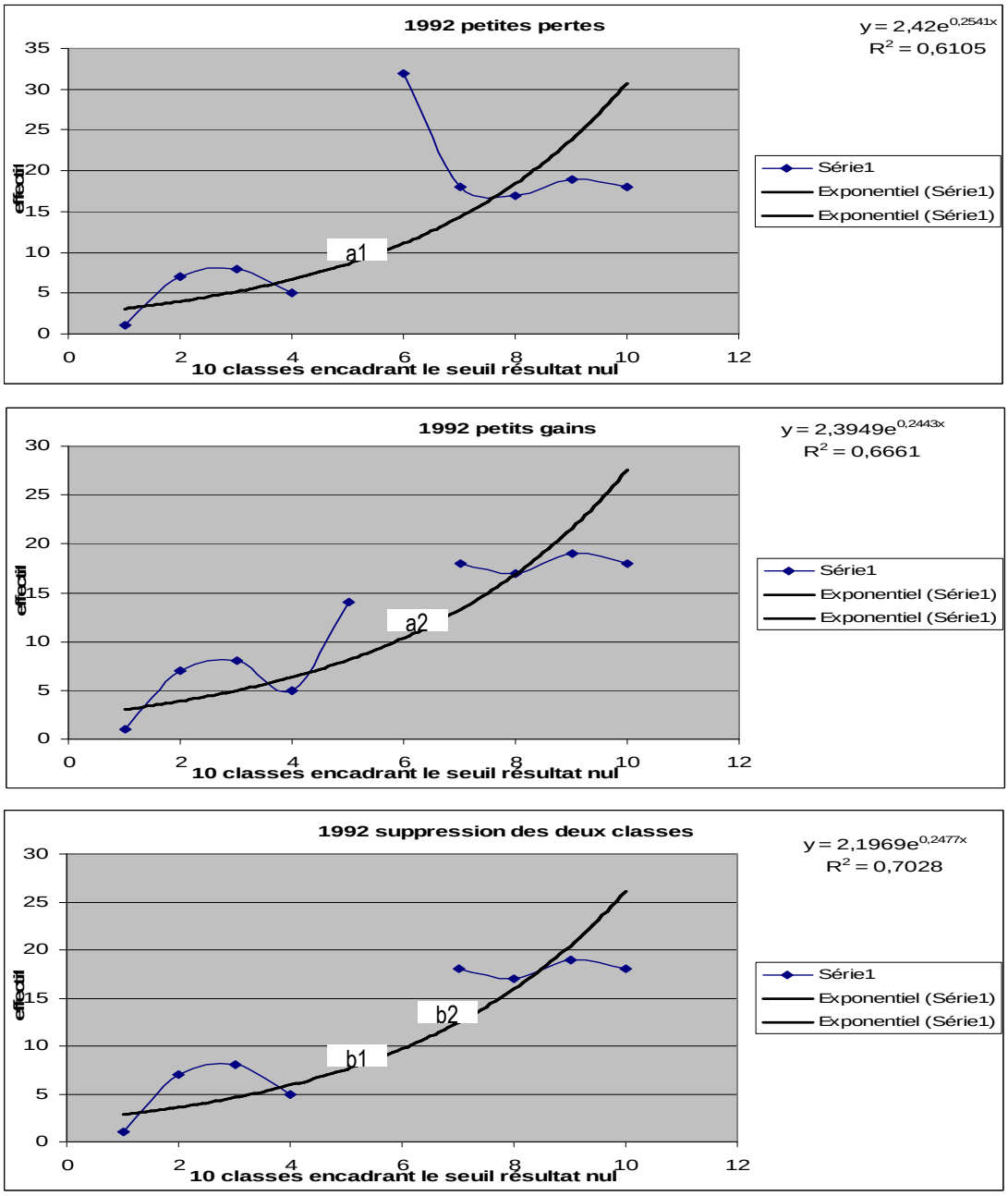
La Figure 42 reprend les mêmes estimations calculées à partir d'une interpolation logarithmique.

Figure 42 : Illustration graphique des interpolations logarithmiques menées sur la seule année 1992



Enfin, la Figure 43 représente les interpolations exponentielles.

Figure 43 : Illustration graphique des interpolations exponentielles menées sur la seule année 1992



Ces 9 interpolations sont récapitulées dans le Tableau 20. Elles permettent d'effectuer 12 estimations pour la seule année 1992. La même démarche a été réalisée sur les 13 années (de 1992 à 2004).

Tableau 20 : Équations des régressions permettant d'estimer les effectifs par interpolation pour la seule année 1992

année1992 interpolation linéaire	Équation	R^2 expliqué
petites pertes uniquement	y = 2.0622x + 2.4324	R2 = 0.4824
petits gains uniquement	y = 1.9392x + 1.3311	R2 = 0.8568
les deux intervalles présumés irrégulières	y = 1.9573x + 0.8598	**R2 = 0.8828**
interpolation exponentielle	**Équation**	**R^2 expliqué**
petites pertes uniquement	y = 2.42e0.2541x	R2 = 0.6105
petits gains uniquement	y = 2.3949e0.2443x	R2 = 0.6661
les deux intervalles présumés irrégulières	y = 2.1969e0.2477x	**R2 = 0.7028**
interpolation logarithmique	**Équation**	**R^2 expliqué**
petites pertes uniquement	y = 9.1293Ln(x) + 0.2	R2 = 0.5548
petits gains uniquement	y = 8.0882Ln(x) - 0.075	R2 = 0.8607
les deux intervalles présumés irrégulières	y = 8.0554Ln(x) - 0.1593	**R2 = 0.8623**

Le Tableau 21 présente la synthèse des travaux réalisés sur les 13 années. Pour ne pas alourdir la lecture, seuls les R^2 des (9 x 13 =) 117 régressions sont présentés. Les équations complètes figurent en Annexe 2 : Équations de régressions.

Tableau 21 : Synthèse des R^2 des (9 x 13 =) 117 régressions

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Moyenne
Interpolation linéaire														
Suppression des petites pertes uniquement	0.4824	0.1887	0.3093	0.5686	0.7116	0.6398	0.634	0.7494	0.6548	0.7071	0.519	0.6714	0.9447	0,599
Suppression des petits gains uniquement	0.8568	0.4891	0.693	0.9029	0.8722	0.8821	0.7062	0.7376	0.6298	0.7694	0.5539	0.6697	0.8521	0,740
les deux intervalles présumés irrégulières	**0.8828**	**0.528**	**0.6958**	**0.9035**	**0.8808**	**0.89**	**0.7275**	**0.7896**	**0.6592**	**0.7841**	**0.5749**	**0.7066**	**0.9535**	0,767
Interpolation exponentielle														
Suppression des petites pertes uniquement	0.6105	0.3631	0.5254	0.7239	0.763	0.7653	0.7093	0.7169	0.7239	0.7803	0.6791	0.6762	0.8785	0,686
Suppression des petits gains uniquement	0.6661	0.5127	**0.7209**	0.8805	0.8429	0.8288	0.7735	0.7135	0.7178	0.8543	0.7217	0.6922	0.7585	0,745
les deux intervalles présumés irrégulières	**0.7028**	**0.5865**	0.7206	**0.8877**	**0.844**	**0.8648**	**0.7743**	**0.7599**	**0.7484**	**0.8554**	**0.7379**	**0.7098**	**0.902**	0,776
Interpolation logarithmique														
Suppression des petites pertes uniquement	0.5548	0.2335	0.35	0.5631	0.7503	0.6812	0.6318	0.6682	**0.586**	0.6549	0.5924	0.6667	**0.8636**	0,600
Suppression des petits gains uniquement	0.8607	0.383	0.5892	0.7508	0.8094	**0.8456**	0.6062	0.5716	0.5065	0.6283	0.5419	0.5819	0.6957	0,644
les deux intervalles présumés irrégulières	**0.8623**	**0.4672**	**0.616**	**0.7686**	**0.8512**	**0.8456**	**0.6686**	**0.6706**	0.5782	**0.6773**	**0.6111**	**0.6693**	0.8609	0,704

Ce tableau appelle deux remarques :

Il apparaît que les interpolations faites en ne supprimant que l'effectif des petites pertes sont moins bonnes que celles qui suppriment les petits gains. Cela signifie que l'irrégularité « supérieure à zéro » est plus importante que l'irrégularité « inférieure à zéro » puisque les effectifs des entreprises faiblement bénéficiaires sont presque toujours plus éloignés de la tendance que les effectifs des entreprises faiblement déficitaires.

Enfin, il apparaît que les R^2 sont toujours supérieurs (à trois exceptions près, entourées dans le tableau) lorsque l'on supprime les deux intervalles entourant l'irrégularité. Cela confirme l'hypothèse de double irrégularité dans la mesure où l'intervalle laissé dans la régression ne renforce pas la tendance, mais au contraire, augmente la variance.

Pour de futures recherches, il semble important de souligner que des estimations par interpolation doivent être toujours menées en supprimant les deux intervalles irréguliers. Par la suite de l'étude, seules les interpolations effectuées sur cette base sont comparées aux autres mesures non paramétriques[86].

Tableau 22 : Tableau des effectifs attendus par interpolation

	Interpolation linéaire				**Interpolation exponentielle**		**Interpolation logarithmique**	
	Suppression du seul intervalle estimé		suppression des deux intervalles irréguliers					
	effectif gauche	***effectif droite***	**effectif gauche**	**effectif droite**	**effectif gauche**	**effectif droite**	**effectif gauche**	**Effectif droite**
1992	*12.74324*	*12.96622*	10.64634	12.60366	7.5802583	9.7108835	12.805366	14.274039
1993	*13.42568*	*11.29054*	10.60366	11.89634	9.3664232	10.544783	11.992088	12.915274
1994	*14.96622*	*13.73649*	12.04268	13.95732	9.8905722	11.715139	14.099220	15.466942
1995	*16.62838*	*16.85811*	14.03049	16.96951	10.710751	13.363779	17.154035	19.212080
1996	*21.62838*	*22.87838*	19.39634	23.35366	13.709891	17.679708	23.748984	26.702411
1997	*20.22973*	*21.68243*	17.45122	21.29878	11.661044	15.498662	21.663191	24.510507
1998	*21.76351*	*22.97973*	19.39634	24.10366	10.923075	14.994411	24.503347	27.929352
1999	*21.41892*	*22.91892*	20.2561	24.2439	15.231726	18.824901	24.492028	27.282095
2000	*24.11486*	*26.7973*	23.39634	28.35366	16.258056	20.550567	28.707153	32.231794
2001	*22.35811*	*24.16892*	20.51829	24.98171	14.733867	18.711695	25.280930	28.430353
2002	*23.39865*	*24.02703*	21.57927	25.17073	16.348435	20.106280	25.633823	28.445039
2003	*20.94595*	*22.27703*	19.80488	23.44512	14.402051	18.034212	23.786437	26.476227
2004	*18.10135*	*20.26351*	17.65244	21.84756	11.595379	15.578732	22.182659	25.209015

Le Tableau 22 récapitule les effectifs calculés à partir des régressions estimées. Chaque année, 12 calculs sont effectués, mais comme par la suite les mesures faites en ne supprimant que les effectifs de l'intervalle estimé sont ignorées, elles n'ont pas été

[86] C'est pourquoi les tableaux de résultats - Tableau 27 et Tableau 28- ne présentent que 8 mesures alors que 11 ont été effectuées.

reportées. Pour mémoire, ces mesures ne sont reportées que pour l'interpolation linéaire (deux premières colonnes du Tableau 22).

Mesures par extrapolation

Les années où les intervalles estimés se situent au sommet de la courbe, c'est-à-dire de 2001 à 2004 uniquement, une estimation par extrapolation est calculée. En effet, ces années-là, une extrapolation peut sembler mieux adaptée qu'une interpolation comme cela a été expliqué dans la section A du chapitre IV[87]. La première étape consiste à déterminer si l'extrapolation doit se faire à partir de la partie gauche ou droite de la courbe.

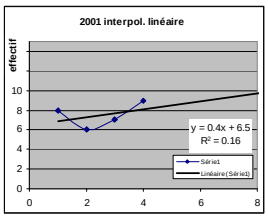
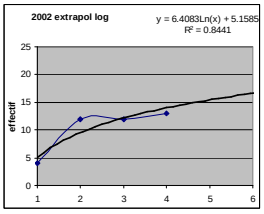
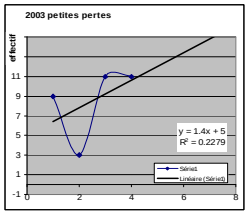
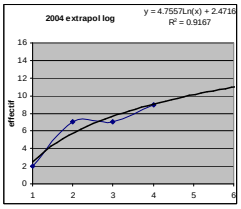
Tableau 23 : Intervalles de référence pour les mesures par extrapolation

	2001	2002	2003	2004
Moyenne	0.00135	0.00123	-0.00020	0.01778
médiane	0.02143	0.01717	0.01845	0.02592
Observation	L'intervalle des entreprises faiblement déficitaires [-0,005 ; 0[se trouve **à gauche de la moyenne et de la médiane**. L'intervalle des entreprises faiblement bénéficiaires [0 ; 0,005 [englobe la moyenne et est à gauche de la médiane.		La moyenne se situe à l'intérieur de l'intervalle des entreprises faiblement déficitaires. Mais les deux intervalles étudiés se situent **à gauche de la médiane**, indicateur moins sensible aux valeurs extrêmes. C'est pourquoi l'extrapolation est faite à partir de la partie gauche de la courbe.	Les intervalles des entreprises faiblement déficitaires et des entreprises faiblement bénéficiaires se trouvent toutes deux **à gauche de la moyenne et de la médiane**.
Intervalles de référence	Dans tous les cas, l'extrapolation se fait à partir de la partie gauche de la courbe.			

En 2003, si la situation par rapport à la moyenne avait été privilégiée, il eut fallu extrapoler l'effectif des entreprises faiblement bénéficiaires à partir de la partie droite de la courbe. La décision d'extrapoler à partir de la partie gauche, justifiée par la moindre volatilité de la médiane, présente également le mérite de conserver toujours la même démarche (toutes les extrapolations sont faites à partir de la partie gauche de la courbe). En conclusion, les extrapolations sont toujours faites à partir des quatre intervalles situés à gauche des intervalles à estimer.

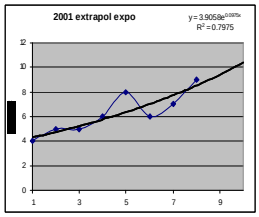
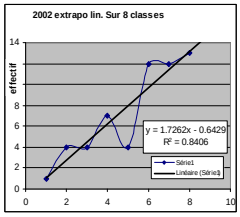
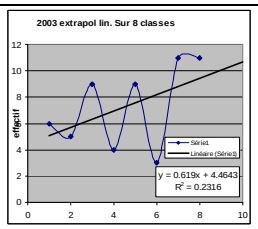
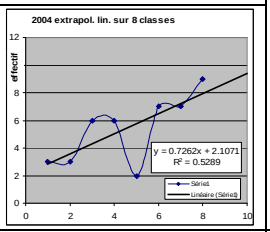
[87] Voir le paragraphe b.1- Les extrapolations linéaires, page 97.

Tableau 24 : Synthèse des extrapolations sur 4 intervalles

	2001	2002	2003	2004	Moyenne des R²
Illustration graphique					
extrapolation logaritm.	y = 0.4523Ln(x) + 7.1406 R² = 0.0444	y = 6.4083Ln(x) + 5.1585 **R² = 0.8441**	y = 2.2136Ln(x) + 6.7413 R² = 0.1235	y = 4.7557Ln(x) + 2.4716 **R² = 0.9167**	**0,4822**
extrapolation exponent.	y = 6.532e0.0507x R² = 0.1408	y = 3.8431e0.3536x R² = 0.6553	y = 4.7001e0.1901x R² = 0.155	y = 1.7638e0.4512x R² = 0.7367	0,4220
extrapolation linéaire	y = 0.4x + 6.5 **R² = 0.16**	y = 2.7x + 3.5 R² = 0.691	y = 1.4x + 5 **R² = 0.2279**	y = 2.1x + 1 R² = 0.8243	0,4758

Hormis en 2002 et 2004, les extrapolations effectuées ont un pouvoir explicatif très faible. Cela peut s'expliquer par le petit nombre (4) d'intervalles utilisés en référence d'une part, et au fait que la volatilité de la distribution semble toujours plus élevée sur la partie gauche de la courbe (puisque les effectifs sont moins nombreux). Pour contrôler les résultats, les calculs ont été refaits en étendant la zone de référence à 8 intervalles.

Tableau 25 : Synthèse des extrapolations sur 8 intervalles

	2001	2002	2003	2004	Moyenne des R²
Illustration graphique					
extrapolation logarithm.	y = 2.0097Ln(x) + 3.586 R2 = 0.7173	y = 5.6696Ln(x) - 0.3905 R2 = 0.7478	y = 1.7769Ln(x) + 4.8945 R2 = 0.1573	y = 2.3349Ln(x) + 2.2799 R2 = 0.4509	0,5183
extrapolation exponent.	y = 3.9058e0.0975x **R2 = 0.7975**	y = 1.3529e0.3117x R2 = 0.7809	y = 4.8509e0.0679x R2 = 0.1156	y = 2.6261e0.1344x R2 = 0.3818	0,5190
extrapolation linéaire	y = 0.5952x + 3.5714 R2 = 0.7631	y = 1.7262x - 0.6429 **R2 = 0.8406**	y = 0.619x + 4.4643 **R2 = 0.2316**	y = 0.7262x + 2.1071 **R2 = 0.5289**	**0,5911**

En prenant 8 intervalles de référence, le pouvoir estimatif des modèles de 2001 et 2002 s'améliore très nettement, mais les estimations de 2004 se détériorent, et celles de 2003

ne s'améliorent pas. Il est également intéressant de constater que sur 4 intervalles de référence, l'extrapolation logarithmique a le meilleur pouvoir estimatif, alors qu'en élargissant la zone de référence, c'est l'extrapolation linéaire qui devient la plus performante. Ce constat semble cohérent avec le fait que la distribution, dont on ne connaît pas la loi, n'est manifestement pas logarithmique. L'approximation logarithmique ne peut être retenue que si l'on se situe au sommet de la distribution, en postulant qu'en ce point extrême, la pente de la courbe tend vers zéro avant de s'inverser. Dès lors que l'on étend la zone de référence, on rencontre la difficulté d'estimer l'allure plus générale de la courbe.

Le Tableau 26 présente les estimations des effectifs attendus à droite et à gauche du seuil du résultat nul sur la base des deux extrapolations : logarithmique sur 4 intervalles, et linéaire sur 8 intervalles de référence.

Tableau 26 : Estimations selon les mesures par extrapolation

	Extrapolation logarithmique sur 4 intervalles		Extrapolation linéaire sur 8 intervalles	
	intervalle [-0,005 ; 0[	**Intervalle [0 ; 0,005[**	**intervalle [-0,005 ; 0[**	**Intervalle [0 ; 0,005[**
2001	7.8685	7.9510	8.9282	9.5234
2002	15.4722	16.6406	14.8929	16.6191
2003	10.3039	10.7075	10.0353	10.6543
2004	10.1256	10.9926	8.6429	9.3691

À ce stade de l'étude, les estimations réalisées à partir de l'extrapolation exponentielle n'ont pas été retenues du fait du (très relatif) moins bon pouvoir explicatif des régressions. Ceci étant dit, c'est avant tout une réflexion sur l'allure attendue de la courbe qui peut permettre de trancher, et l'allure exponentielle n'est pas à écarter. Ce sujet est abordé dans la partie 3 de la thèse.

c. Résultats ; comparaison des différentes mesures

Toutes les mesures effectuées ont été récapitulées dans le Tableau 27 en ce qui concerne les irrégularités à gauche du seuil (intervalle [-0,005 ; 0[) et dans le Tableau 28 en ce qui concerne les irrégularités à droite du seuil (intervalle [0 ; 0,005[). En ce qui concerne les interpolations, seules celles faites en supprimant les deux intervalles supposés irréguliers de la zone de référence ont été conservées (trois mesures à droite et trois mesures à gauche, soit six mesures au lieu des douze présentées page 141). En ce

qui concerne les extrapolations, elles n'ont été faites de sur la période 2001 à 2004 ; une partie du tableau est donc blanche. Les irrégularités sont exprimées en nombre d'entreprises en surplus ou manquante (différence entre le nombre d'entreprises observées sur l'intervalle moins le nombre attendu par estimation).

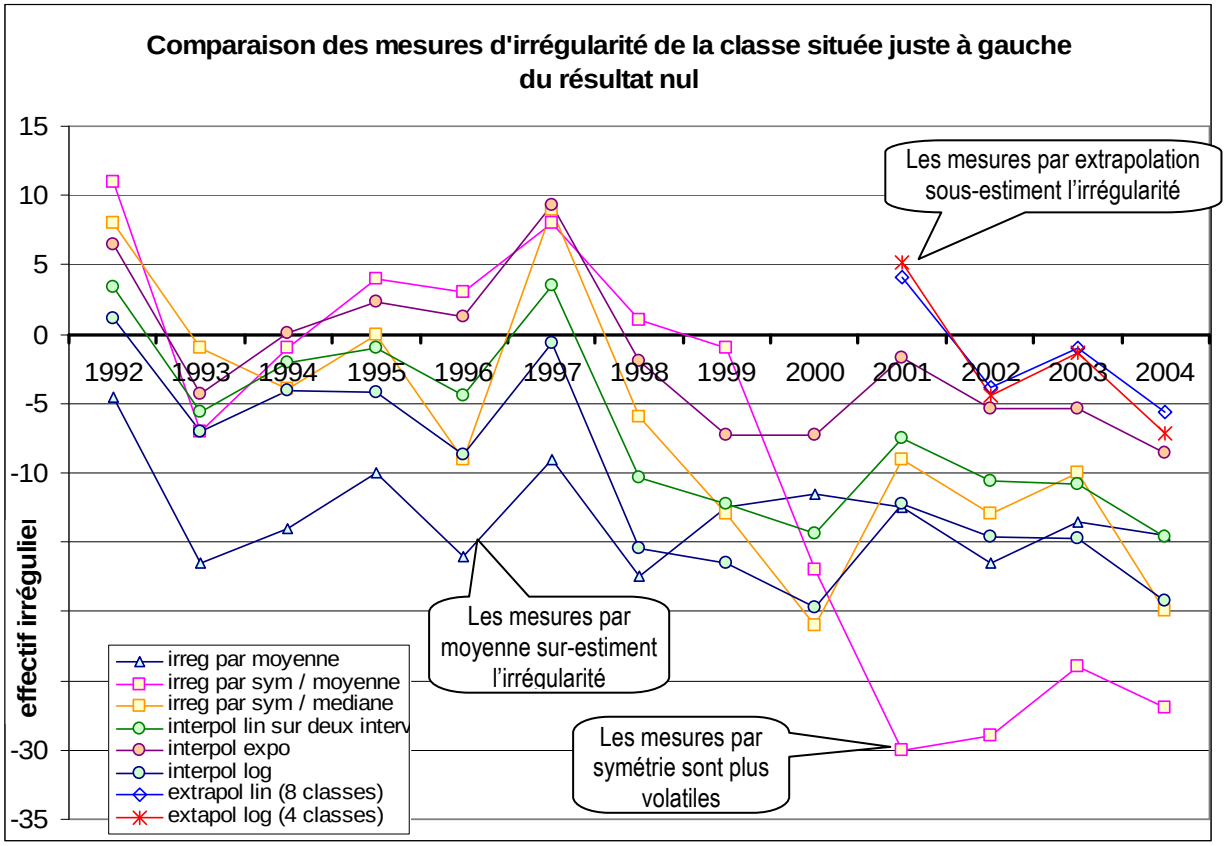
c.1. Mesures de l'irrégularité gauche

Tableau 27 : Mesures non paramétriques de l'irrégularité sur [-0,005 ; 0[

Famille	Moyenne	Symétrie		Extrapolation		Interpolation		
Année	Irrégularité mesurée par moyenne des intervalles adjacents	Irrégularité mesurée par symétrie par rapport à la moyenne	Irrégularité mesurée par symétrie par rapport à la médiane	Irrégularité mesurée par extrapolation logarithmique sur 4 intervalles	Irrégularité mesurée par extrapolation linéaire sur 8 intervalles	Irrégularité mesurée par interpolation linéaire (exclusion de 2 intervalles)	Irrégularité mesurée par interpolation exponentielle (exclusion de 2 intervalles)	Irrégularité mesurée par interpolation logarithmique (exclusion de 2 intervalles)
1992	-4.5	11	8			3.3536	6.4197	1.1946
1993	-16.5	-7	-1			-5.6036	-4.3664	-6.9920
1994	-14	-1	-4			-2.0426	0.1094	-4.0992
1995	-10	4	0			-1.0304	2.2892	-4.1540
1996	-16	3	-9			-4.3963	1.2901	-8.7489
1997	-9	8	9			3.5487	9.3389	-0.6631
1998	-17.5	1	-6			-10.3963	-1.9230	-15.5033
1999	-12.5	-1	-13			-12.2561	-7.2317	-16.4920
2000	-11.5	-17	-21			-14.3963	-7.2580	-19.7071
2001	-12.5	-30	-9	5.1314	4.0718	-7.5182	-1.7338	-12.2809
2002	-16.5	-29	-13	-4.4722	-3.8929	-10.5792	-5.3484	-14.6338
2003	-13.5	-24	-10	-1.3039	-1.0353	-10.8048	-5.4020	-14.7864
2004	-14.5	-27	-20	-7.1256	-5.6429	-14.6524	-8.5953	-19.1826

Pour comparer les 8 séries de mesures effectuées, les données de ce tableau sont représentées graphiquement Figure 44. C'est ce graphique qui est commenté par la suite.

Figure 44 : Représentation graphique des irrégularités mesurées sur [-0,005 ; 0[



Il apparaît graphiquement (1) d'importants écarts entre les différentes mesures, mais également (2) une certaine cohérence dans leur évolution. Les mesures par symétries (surtout par rapport à la moyenne) sont plus volatiles. La mesure par moyenne des deux intervalles adjacents tend à majorer l'irrégularité. Les extrapolations semblent, quant à elles, minorer l'irrégularité. De manière générale, toutes les mesures révèlent des irrégularités essentiellement négatives (sous-représentation des entreprises faiblement bénéficiaires). Ce constat va dans le sens des travaux antérieurs. L'amplitude de ces irrégularités tend à augmenter au cours des 13 années de l'étude, ce qui n'avait jamais été observé auparavant en France.

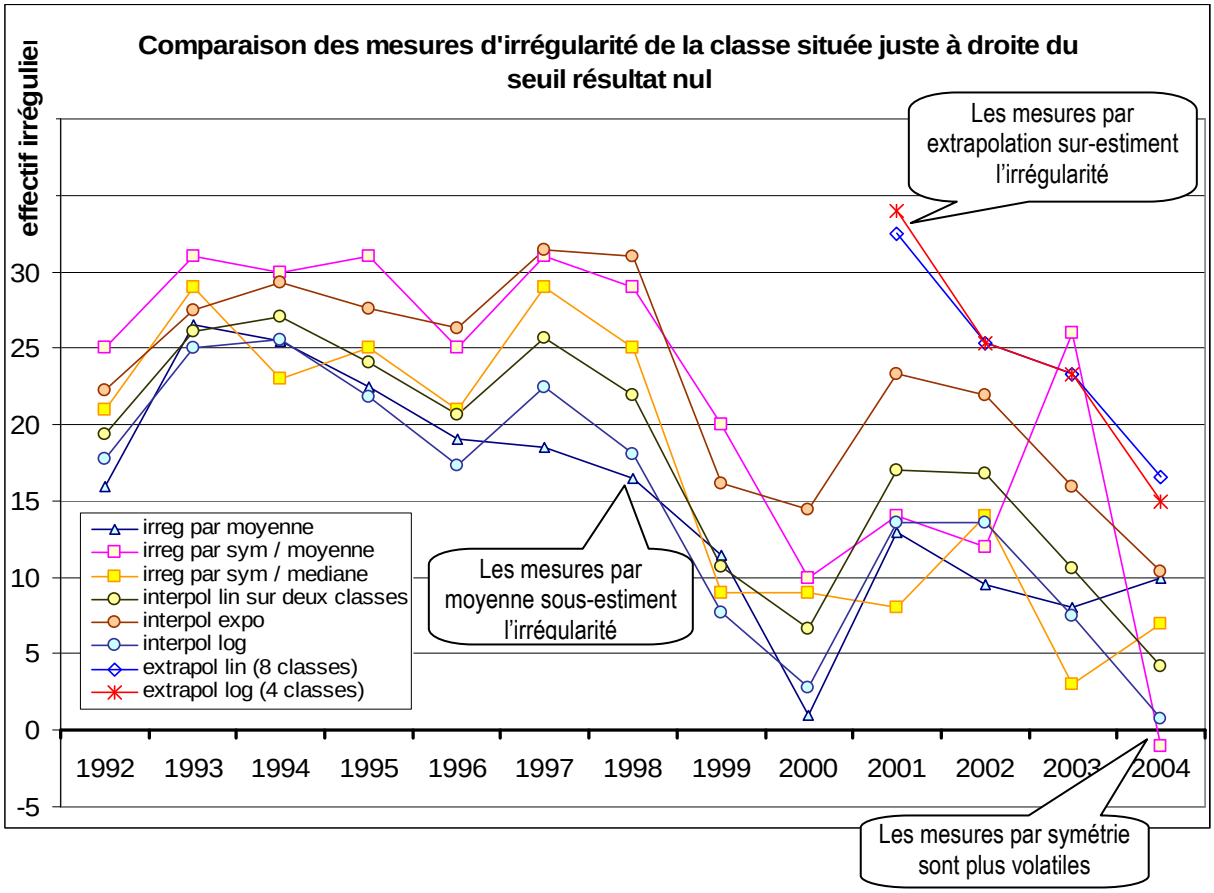
c.1. Mesures de l'irrégularité droite

Tableau 28 : Mesures non paramétriques de l'irrégularité droite sur [0 ; 0,005[

Famille	Moyenne	Symétrie		Extrapolation		Interpolation		
Année	Irrégularité mesurée par moyenne des intervalles adjacents	Irrégularité mesurée par symétrie par rapport à la moyenne	Irrégularité mesurée par symétrie par rapport à la médiane	Irrégularité mesurée par extrapolation logarithmique sur 4 intervalles	Irrégularité mesurée par extrapolation linéaire sur 8 intervalles	Irrégularité mesurée par interpolation linéaire	Irrégularité mesurée par interpolation exponentielle	Irrégularité mesurée par interpolation logarithmique
1992	16	25	21			19.3963	22.2891	17.7259
1993	26.5	31	29			26.1036	27.4552	25.0847
1994	25.5	30	23			27.0426	29.2848	25.5330
1995	22.5	31	25			24.0304	27.6362	21.7879
1996	19	25	21			20.6463	26.3202	17.2975
1997	18.5	31	29			25.7012	31.5013	22.4894
1998	16.5	29	25			21.8963	31.0055	18.0706
1999	11.5	20	9			10.7561	16.1750	7.7179
2000	1	10	9			6.6463	14.4494	2.7682
2001	13	14	8	34.0489	32.4766	17.0182	23.2883	13.5696
2002	9.5	12	14	25.3593	25.3809	16.8292	21.8937	13.5549
2003	8	26	3	23.2924	23.3457	10.5548	15.9657	7.5237
2004	10	-1	7	15.0073	16.6309	4.15244	10.4212	0.7909

Comme précédemment, la comparaison des 8 séries de mesures est faite à partir de leur représentation graphique de la Figure 45.

Figure 45 : Représentation graphique des irrégularités droites mesurées sur [0 ; 0,005[



À nouveau, il apparaît (1) d'importants écarts entre les différentes mesures, tout en constatant (2) une certaine cohérence dans leur évolution. Les mesures par symétrie semblent encore une fois plus volatiles. Les extrapolations conduisent cette fois à une majoration de l'irrégularité. De manière générale, toutes les mesures révèlent des irrégularités positives (sur-représentation des entreprises faiblement bénéficiaires) ce qui est conforme aux résultats antérieurs. Leur amplitude tend à diminuer au cours des 12 années de l'étude, ce qui n'avait jamais été encore observé.

d. Hiérarchisation des mesures

Les irrégularités sont mesurées selon 11 méthodes différentes et les résultats présentent des différences importantes. Les méthodes par interpolation ne retirant qu'un intervalle ont déjà été écartées au profit des interpolations fondées sur le retrait des deux intervalles encadrant le seuil. Pour départager les 8 autres méthodes, cinq critères sont étudiés.

d.1. La volatilité des mesures

La volatilité des mesures d'effectifs théoriques effectuées sur les 13 années est mesurée avec chaque méthode. Une volatilité élevée est interprétée comme un indicateur de fragilité de la méthode.

Tableau 29 : Volatilité comparée des différentes mesures non paramétriques

Méthode d'estimation de l'effectif à gauche du seuil	**Écart type** (classement par ordre croissant)	**Méthode d'estimation de l'effectif à droite du seuil**	**Écart type** (classement par ordre croissant)
interpol expo	2.80638466	extrapol lin sur 8 intervalles	3.43322953
extrapol lin sur 8 intervalles	2.90806734	interpol expo	3.59825199
extrapol log	3.21677757	extrapol log	3.64625346
moyenne	4.19630485	interpol lin	5.31491673
interpol lin	4.2520116	interpol log	6.19856836
interpol log	5.40692247	moyenne	7.28297101
symétrie / médiane	7.29769897	symétrie / médiane	7.41187543
symétrie / MOYenne	13.4092391	symétrie / MOYenne	8.3112528

Confirmant l'analyse graphique des Figure 44 et Figure 45, il apparaît que les mesures par symétrie sont les plus volatiles. Les limites de ces méthodes ont déjà été soulignées dans le chapitre IV. Ce résultat confirme l'idée que les mesures par symétrie sont peu fiables. La moyenne étant plus sensible aux valeurs extrêmes, c'est sans surprise que la symétrie par rapport à la moyenne apparaît comme étant l'indicateur le moins stable.

Les mesures par moyenne arithmétique sont également peu stables, surtout en ce qui concerne la mesure des effectifs à droite du seuil. Ce résultat est important à souligner dans la mesure où cette méthode est pourtant la plus couramment utilisée.

Les mesures par extrapolation n'ont été effectuées que sur 4 années, et la faible volatilité ne peut être interprétée ici.

Parmi les mesures par interpolation, il apparaît une hiérarchie : les interpolations exponentielles sont moins volatiles, puis viennent les interpolations linéaires, et enfin les interpolations logarithmiques. Ce résultat confirme l'idée que, dans la plupart des cas, les seuils se situent dans une phase fortement croissante de la courbe, mieux ajustée par une fonction exponentielle.

En conclusion, les mesures par symétrie, et dans une moindre mesure par moyenne des intervalles adjacents semblent plus volatiles, ce qui peut être interprété comme un signe de forte dépendance méthodologique de ces mesures vis-à-vis des paramètres de calcul.

Or ce sont les mesures les plus fréquemment utilisées dans la littérature. Les interpolations, exponentielles ou linéaires, apparaissent plus pertinentes.

d.2. La qualité de la régression

Les méthodes sont classées en fonction du R^2 de la fonction de régression utilisée pour prédire les effectifs des intervalles adjacents (présumés non irréguliers) aux intervalles estimés. Ces R^2 ont déjà été calculés dans le cadre des interpolations et extrapolations réalisées. Dans le cas des moyennes des effectifs adjacentes, il semble inutile de les calculer dans la mesure où l'équation de la droite obtenue passant par les deux effectifs encadrant l'effectif estimé ne pourra jamais être meilleure pour représenter les 8 intervalles adjacents, que la droite estimée à partir de ces 8 intervalles. Sur le critère du R^2, la méthode de la moyenne des intervalles adjacents n'est donc jamais meilleure que l'interpolation linéaire.

Tableau 30 : Qualité comparée des différentes régressions

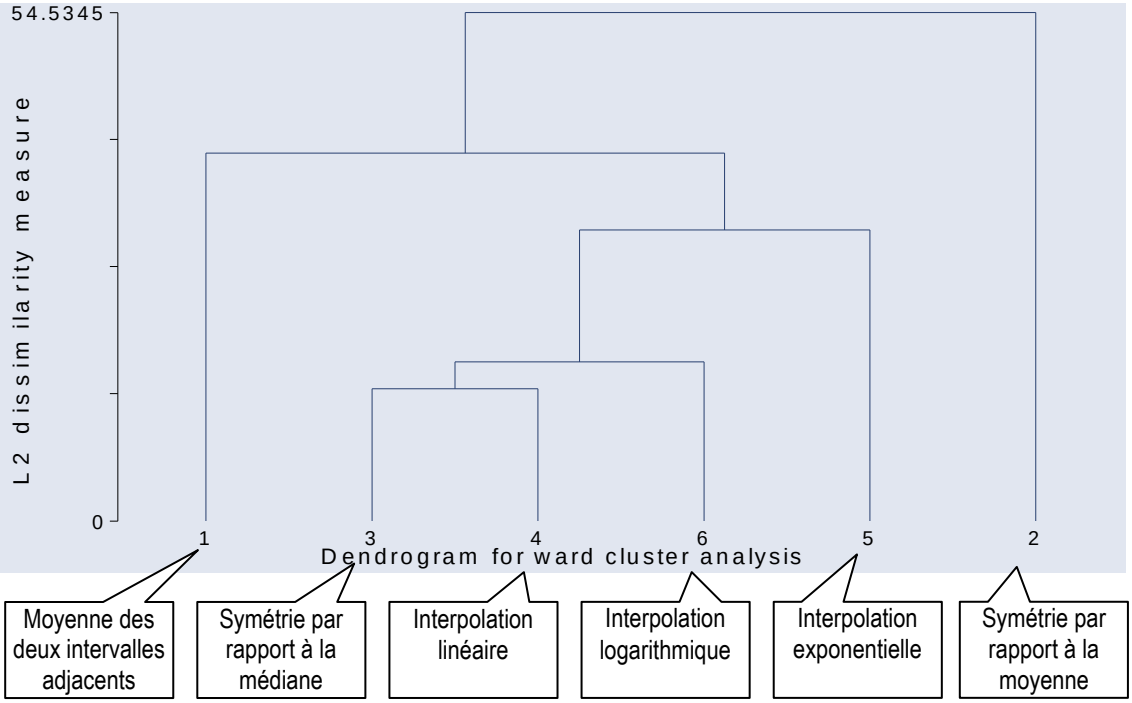
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**Moyenne des R^2** (classement par ordre croissant)	Écart Type
Extrapolation expo										0.140	0.655	0.155	0.736	0.421	0.318
Extrapolation lin										0.16	0.691	0.227	0.824	0.480	0.33
Extrapo log										0.044	0.844	0.123	0.916	0.482	0.461
Interpolation logarithmique	0.862	0.467	0.616	0.768	0.851	0.845	0.668	0.670	0.578	0.677	0.611	0.669	0.860	0.703	0.125
interpolation linéaire	0.882	0.528	0.695	0.903	0.880	0.89	0.727	0.789	0.659	0.784	0.574	0.706	0.953	**0.767**	**0.132**
interpolation exponent.	0.702	0.586	0.720	0.887	0.844	0.864	0.774	0.759	0.748	0.855	0.737	0.709	0.902	**0.776**	0.090

Selon ce critère, les interpolations sont naturellement plus fiables que les extrapolations (qui n'ont pas été faites sur les mêmes intervalles de référence). Ce qui est intéressant de remarquer, c'est que les interpolations exponentielles fournissent en général une meilleure estimation, suivies de près par les interpolations linéaires. Ces dernières, si elles fournissent de meilleurs R^2 certaines années, ont un écart type plus élevé, ce qui signifie que cette méthode peut être mieux adaptée certaines années, mais moins bien d'autres années. Ce résultat peut s'expliquer par le fait que l'irrégularité est située en général sur la partie fortement croissante de la courbe de distribution.

d.3. L'analyse de classification hiérarchique

Les dendrogrammes (ou classifications hiérarchiques) des mesures des irrégularités à gauche (Figure 46), puis à droite (Figure 47), sont réalisés. Les extrapolations sont exclues de cette analyse car elles ont été effectuées sur 4 années uniquement.

Figure 46 : Classification hiérarchique des mesures non paramétriques de l'irrégularité gauche

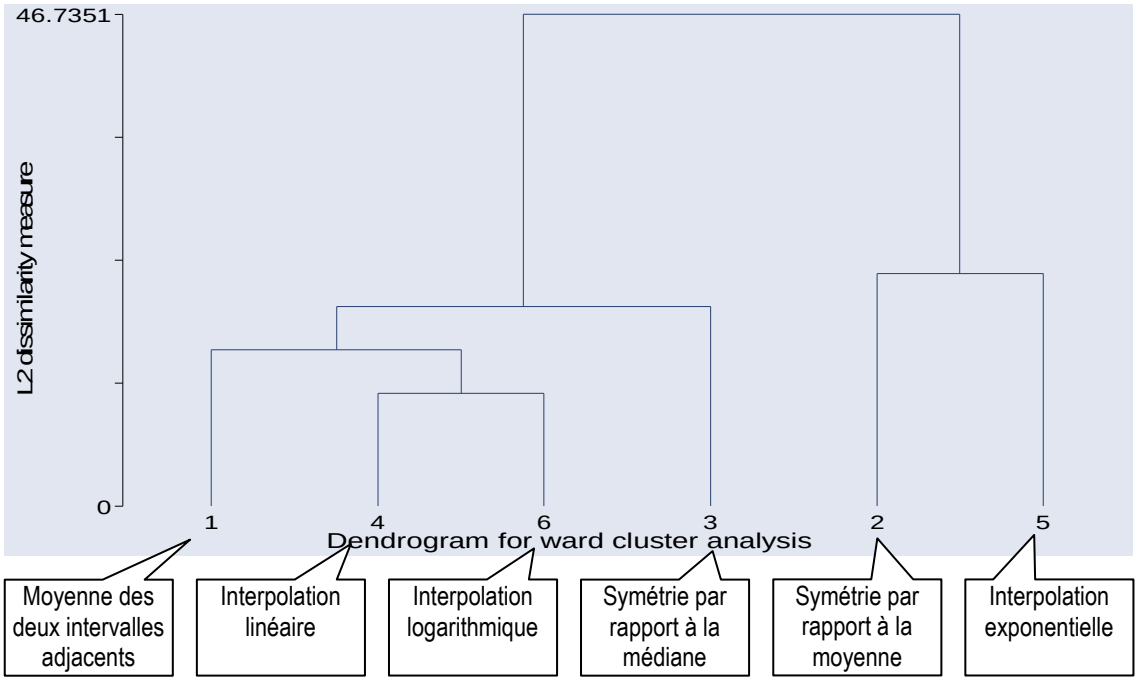


La mesure de l'irrégularité à gauche du seuil par moyenne des deux intervalles adjacents et la mesure par symétrie par rapport à la moyenne sont les plus atypiques. Ces résultats confirment l'analyse graphique de la Figure 44 : la mesure par symétrie par rapport à la moyenne est celle qui a la plus forte pente, alors que la mesure par moyenne est quasiment stable dans le temps.

Étrangement, la mesure par symétrie par rapport à la médiane est assez proche des mesures par interpolation, et notamment de l'interpolation linéaire. Ceci n'est vrai que pour l'irrégularité à gauche du seuil. Sur le graphe de l'évolution des mesures, il apparaît que la mesure par interpolation linéaire tend toujours à fournir une valeur de l'irrégularité plus élevée (ou positive, ou faiblement négative), donc à modérer le sous-effectif à gauche du seuil.

Enfin, l'interpolation exponentielle est la plus atypique. Le dendrogramme des mesures de l'irrégularité à droite est présenté (Figure 47) :

Figure 47 : Classification hiérarchique des mesures non paramétriques de l'irrégularité droite



À nouveau, il apparaît que les mesures par interpolation logarithmique et par interpolation linéaire sont les plus proches. Les mesures par symétrie occupent des positions plus éloignées. La mesure par moyenne des deux intervalles adjacents est cette fois plus proche des interpolations linéaires. Enfin, la mesure par interpolation exponentielle, qui tend à surestimer l'irrégularité (sur pondérer le sureffectif) se retrouve cette fois proche de la mesure par symétrie par rapport à la moyenne.

Ces résultats mettent en évidence les caractéristiques de chaque mesure, caractéristiques liées à l'allure générale des distributions autour du seuil résultat nul. En général, les effectifs juste inférieurs au seuil sont nettement plus faibles que les effectifs juste supérieurs. Cette irrégularité ne se restreint pas au seul intervalle jouxtant le seuil. Les effectifs à gauche étant plus faibles, il en découle que la volatilité des effectifs est plus grande.

Par ailleurs, l'interpolation exponentielle aura tendance à « creuser » l'allure de la courbe, et donc à sous-estimer les effectifs estimés alors que l'interpolation logarithmique aura l'effet inverse. Il apparaît dès lors qu'en général, l'interpolation exponentielle tend à révéler davantage les sureffectifs, ou à minimiser les sous-effectifs, ce qui est l'inverse de l'interpolation logarithmique.

La mesure par moyenne des deux intervalles adjacents tendrait à se rapprocher de l'interpolation linéaire, mais elle est beaucoup plus volatile, d'autant plus que l'un des deux intervalles de référence est nécessairement irrégulier lui-même.

Les mesures par symétrie tendent à prendre comme état de référence des effectifs à droite de la distribution qui sont très volatiles puisqu'il n'y a pas de lissage sur plusieurs intervalles. Ce qui explique une volatilité plus importante. Notons que la symétrie par rapport à la moyenne apparaît toujours comme la plus atypique.

En conclusion, les mesures par symétrie peuvent être considérées comme très atypiques, alors que les mesures par interpolation, notamment linéaire et logarithmique, sont les plus homogènes.

d.4. La comparaison des tendances

Pour chaque série, une régression linéaire est effectuée, cette tendance est tracée. La Figure 48 présente les irrégularités à gauche du seuil.

Figure 48 : Comparaison des tendances des mesures non paramétriques (irrégularité gauche)

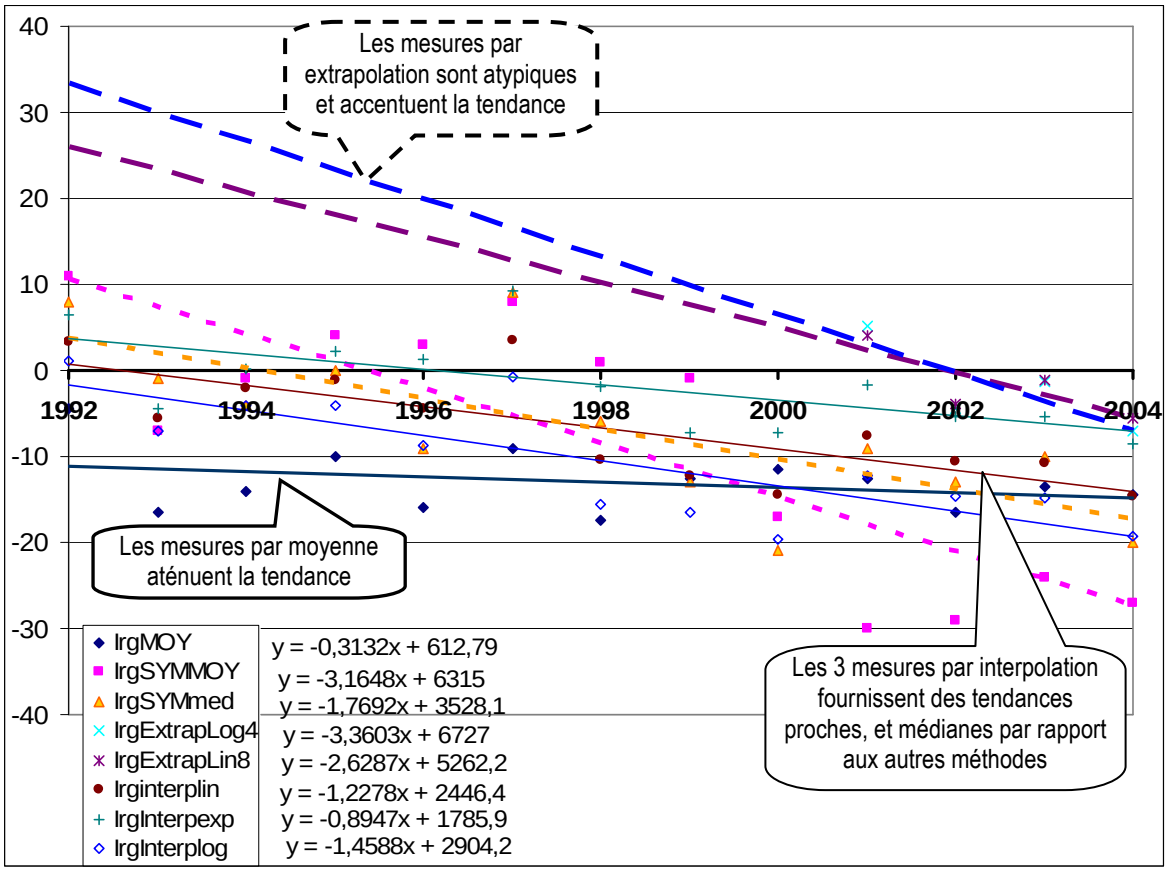


Tableau 31 : Comparaison des coefficients alpha (pente des régressions) ; irrégularité gauche

Type de mesure	Pente de la droite (classement par ordre croissant)
IrgMOY	-0.3132
IrgInterpexp	-0.8947
Irginterplin	-1.2278
IrgInterplog	-1.4588
IrgSYMmed	-1.7692
IrgExtrapLin8	-2.6287
IrgSYMMOY	-3.1648
IrgExtrapLog4	-3.3603

Il est intéressant de souligner que toutes les mesures fournissent une tendance décroissante (pente négative). Toutes les mesures tendent donc à fournir un résultat similaire sur la tendance à moyen terme des irrégularités.

Les méthodes qui fournissent les plus fortes pentes, donc qui tendent à montrer que l'irrégularité (le sous-effectif) à gauche du seuil croît fortement durant la période étudiée, sont les mesures par symétrie et par extrapolation. Les mesures par symétrie ont déjà été identifiées comme étant les plus volatiles, et les mesures par extrapolation n'ont été faites que sur 4 années au lieu de 13. Ces tendances apparaissent donc comme étant plus fragiles.

À l'inverse, la méthode qui tend à limiter la tendance, voire à la masquer, est la mesure par moyenne des deux intervalles adjacents.

Les méthodes par interpolation fournissent des mesures de tendances relativement proches, et médianes par rapport aux autres méthodes. Elles apparaissent donc, d'une certaine façon, plus fiables.

La même démarche a été menée avec les irrégularités à droite du seuil (Figure 49).

Figure 49 : Comparaison des tendances des mesures non paramétriques (irrégularité droite)

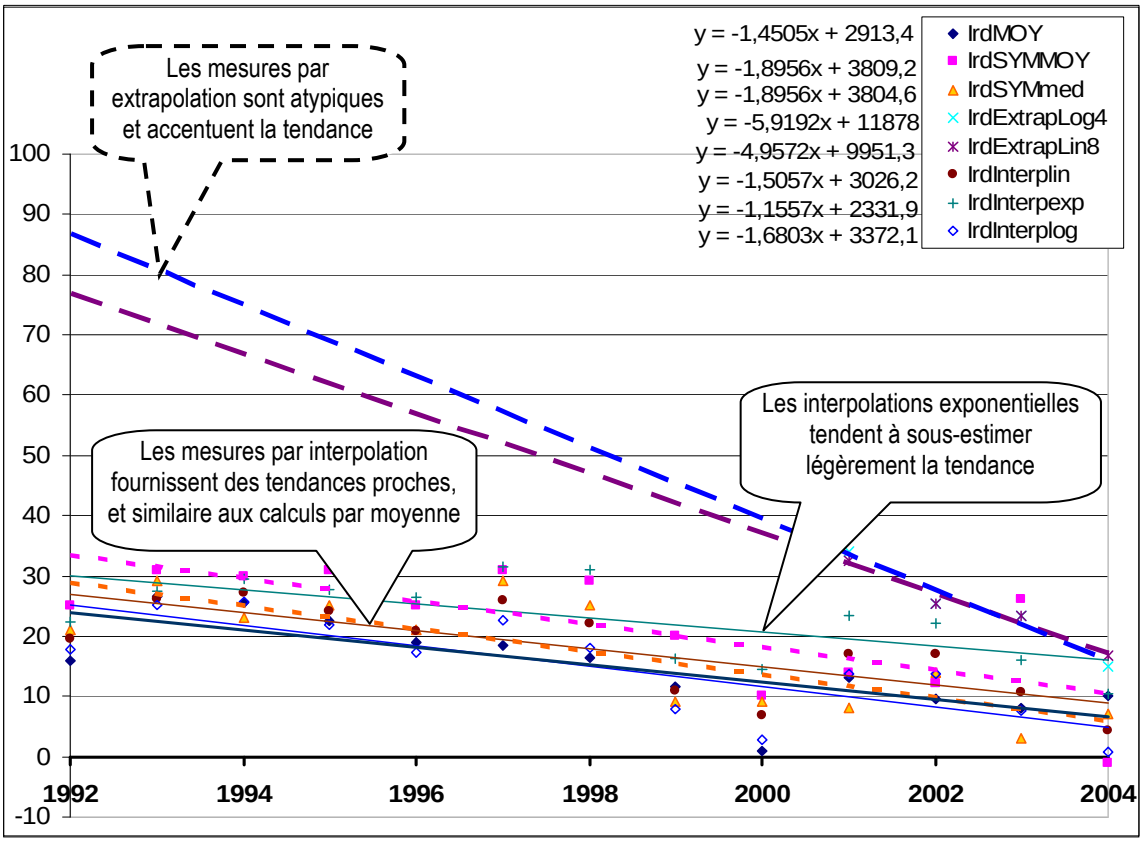


Tableau 32 : Comparaison des coefficients alpha (pente des régressions) ; irrégularité droite

Type de mesure	Pente de la courbe (classement par ordre croissant)
IrdInterpexp	-1.1557
IrdMOY	-1.4505
IrdInterplin	-1.5057
IrdInterplog	-1.6803
IrdSYMMOY	-1.8956
IrdSYMmed	-1.8956
IrdExtrapLin8	-4.9572
IrdExtrapLog4	-5.9192

En ce qui concerne l'irrégularité à droite du seuil, les conclusions sont similaires. Toutes les mesures fournissent une tendance décroissante (pente négative) ce qui signifie que toutes les mesures tendent à fournir la même tendance à moyen terme.

Les méthodes qui fournissent les plus fortes pentes, donc qui tendent à montrer que l'irrégularité (le sur-effectif) à droite du seuil diminue fortement durant la période étudiée, sont les mesures par extrapolation, et dans une moindre mesure par symétrie. Or ce sont les mesures les moins fiables.

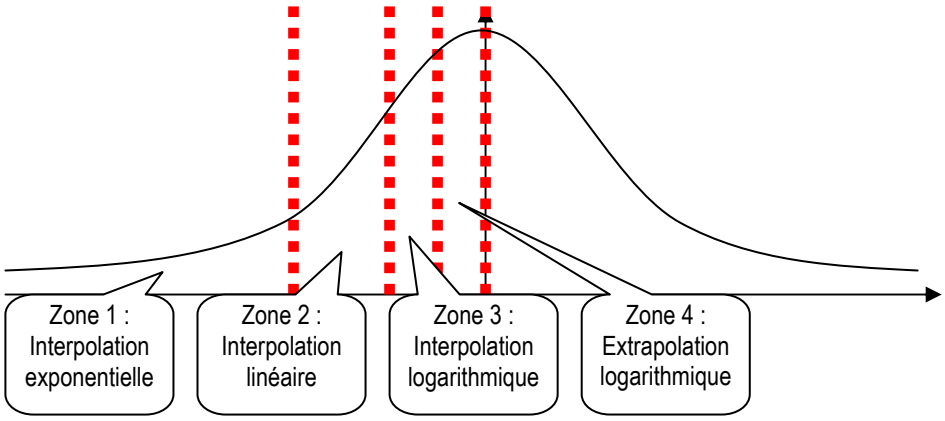
À l'inverse, la méthode par moyenne des deux intervalles adjacents tend à limiter la tendance. Les méthodes par interpolation fournissent des mesures de tendances relativement proches, et médianes par rapport aux autres méthodes, à l'exception près de l'interpolation exponentielle, qui cette fois, tend à limiter la tendance.

En conclusion, le choix de la méthode peut influencer très sensiblement les résultats de l'analyse de la tendance. Certaines méthodes sous-estiment, voire masquent une tendance que d'autres méthodes tendent à amplifier largement. Ceci étant dit, toutes les méthodes fournissent une tendance négative en ce qui concerne les deux irrégularités au cours de la période étudiée, c'est-à-dire que, tenant compte du signe de l'irrégularité, il y a une croissance de l'irrégularité à gauche et une diminution de l'irrégularité à droite. Ce constat n'a encore jamais été réalisé dans la littérature comptable. Cette double évolution tend à montrer que l'étude de chacune des irrégularités est nécessaire pour une meilleure compréhension du phénomène.

d.5. Le choix théorique du meilleur outil

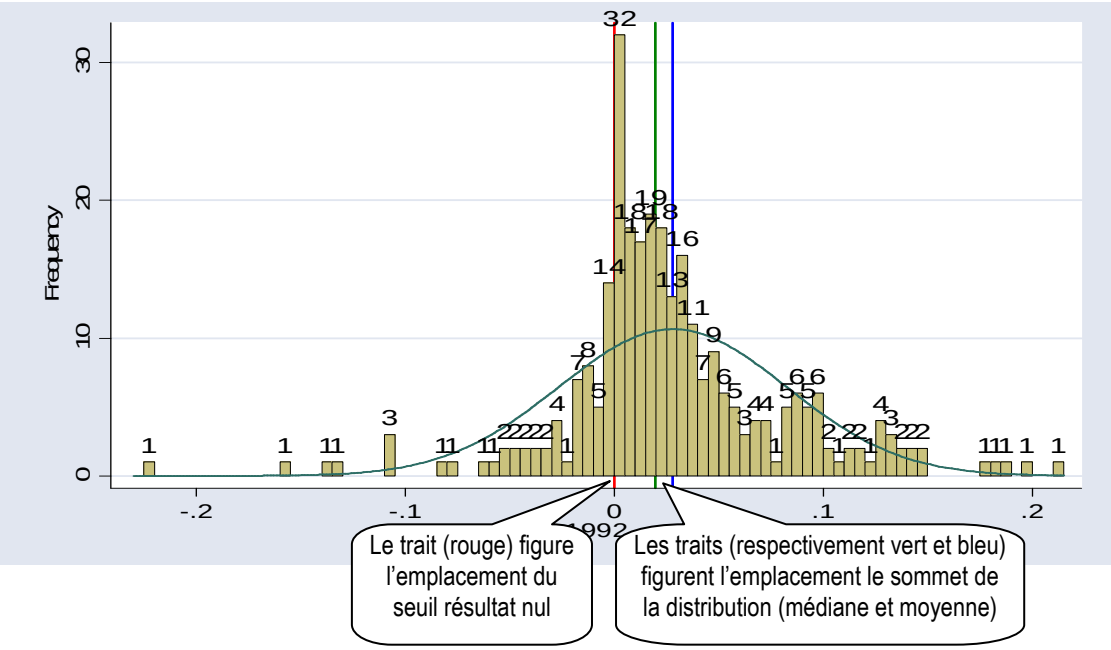
Enfin, l'étude cherche à déterminer la pertinence de la méthode utilisée en fonction de l'emplacement de l'irrégularité dans la distribution. Pour cela, l'emplacement du résultat nul est observé dans la distribution année par année. La Figure 50 rappelle les conclusions de la réflexion menée dans le chapitre IV (voir Figure 35 page 112) pour déterminer la mesure théoriquement la plus adaptée.

Figure 50 : Choix de la mesure appropriée en fonction de la position du seuil sur la distribution



Ainsi, lorsque le seuil se situe dans la partie gauche de la distribution, l'estimation théoriquement la plus adaptée est l'interpolation exponentielle. Si elle se situe au sommet, l'extrapolation logarithmique est plus adaptée. Année après année, les distributions sont donc tracées, et la place du seuil « résultat nul » est indiquée. La Figure 51 illustre cette démarche pour la seule année 1992.

Figure 51 : Position du seuil et choix de la mesure appropriée en 1992[88]



Il apparaît ainsi qu'en 1992, le seuil se situe sur la pente positive décroissante de la distribution (zone 3 de la Figure 50). Autrement dit, la mesure théoriquement adaptée est l'interpolation logarithmique. Cette analyse a été conduite pour chaque année de 1992 à 2004, et le Tableau 33 récapitule les résultats[89].

[88] Sur ce graphe, la loi Normale associée à la distribution est tracée pour faciliter l'analyse graphique. Mais elle ne présume pas de la loi réelle de distribution.

[89] Le détail des analyses graphiques est fourni en Annexe 3 : Emplacement de l'irrégularité dans la distribution annuelle et choix de la méthode de mesure appropriée .

Tableau 33 : Emplacement de l'irrégularité dans la distribution annuelle et choix de la méthode appropriée (hypothèse sous-jacente : la distribution régulière a une forme de cloche)

Année	Position du seuil **« résultat nul »**	Mesure adaptée
1992	Pente positive décroissante	Interpolation logarithmique
1993	Pente positive décroissante	Interpolation logarithmique
1994	Pente positive décroissante	Interpolation logarithmique
1995	Pente positive décroissante	Interpolation logarithmique
1996	Pente positive décroissante	Interpolation logarithmique
1997	Pente positive décroissante	Interpolation logarithmique
1998	Pente positive décroissante	Interpolation logarithmique
1999	Pente positive décroissante	Interpolation logarithmique
2000	Pente positive décroissante	Interpolation logarithmique
2001	Sommet de la courbe	Extrapolation (logarithmique ou linéaire)
2002	Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)
2003	Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)
2004	Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)

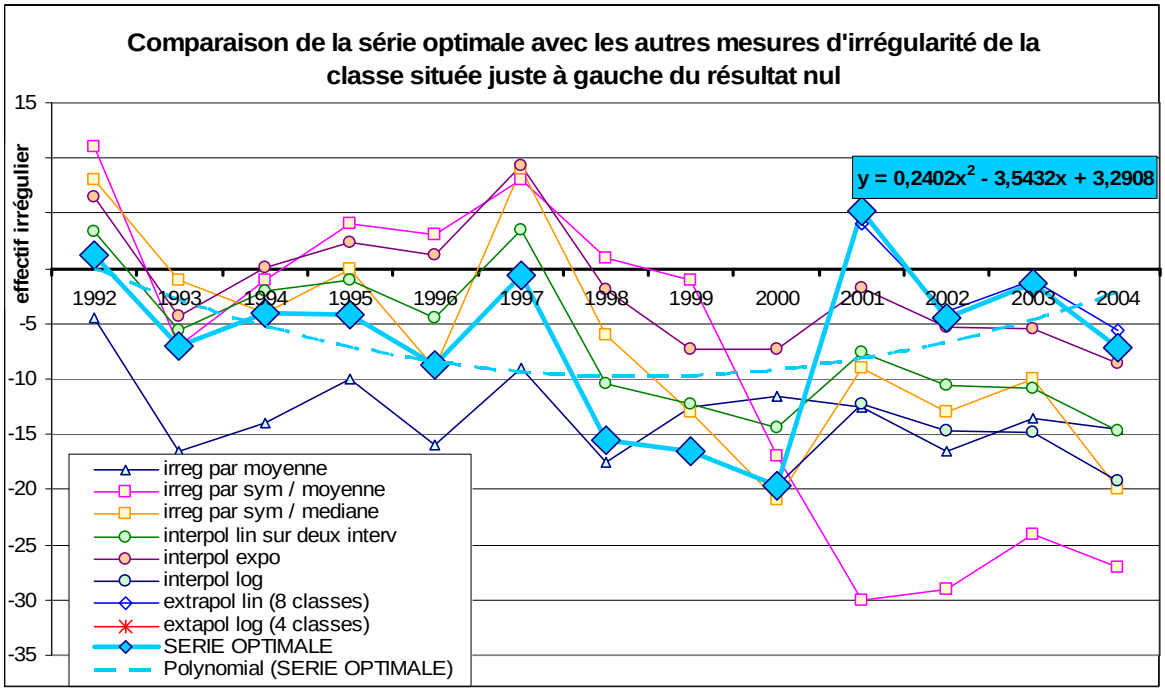
Comme le montre le Tableau 33, l'irrégularité se situe presque toujours dans la phase ascendante proche du sommet de la courbe (zone 3 de la Figure 50). Les mesures optimales sont alors ou les interpolations, ou les extrapolations logarithmiques. Le Tableau 34 présente la série optimale des mesures déterminées par observation des distributions, sous hypothèse que la distribution présente une allure « en cloche ».

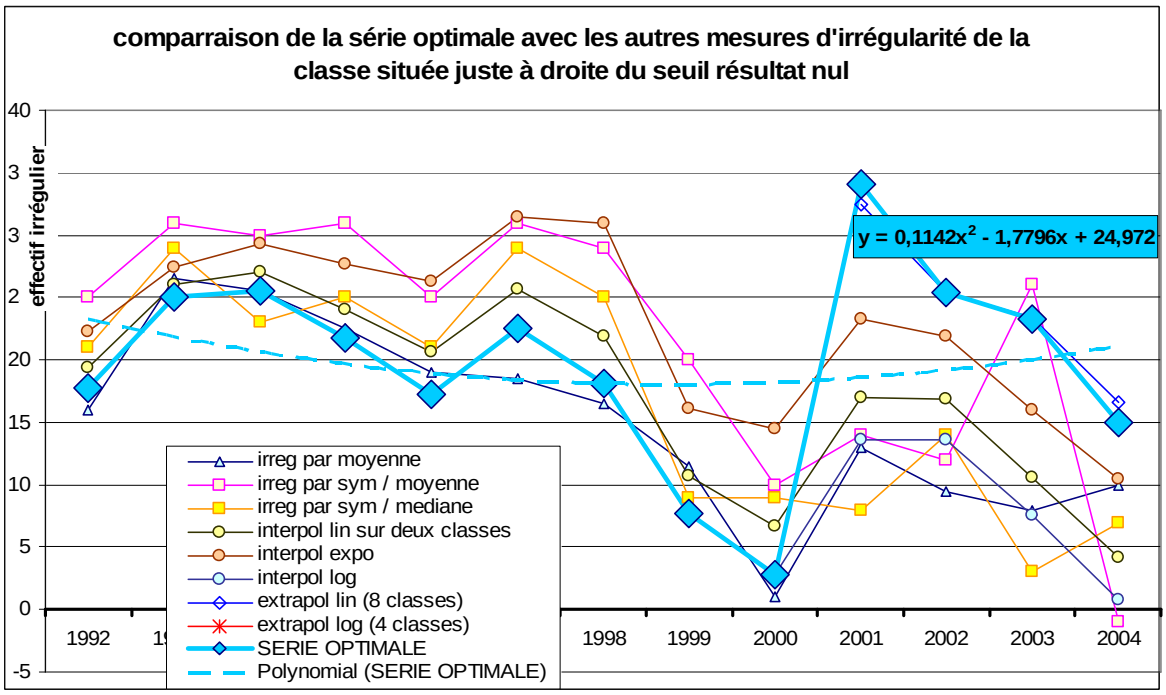
Tableau 34 : Série de mesures optimale si la distribution attendue est en forme de cloche

Année	Méthode privilégiée	Mesure de référence pour l'intervalle à gauche	Mesure de référence pour l'intervalle à droite
1992	Interpolation logarithmique	0.09	1.24
1993	Interpolation logarithmique	-0.58	1.94
1994	Interpolation logarithmique	-0.29	1.65
1995	Interpolation logarithmique	-0.24	1.13
1996	Interpolation logarithmique	-0.37	0.65
1997	Interpolation logarithmique	-0.03	0.92
1998	Interpolation logarithmique	-0.63	0.65
1999	Interpolation logarithmique	-0.67	0.28
2000	Interpolation logarithmique	-0.69	0.09
2001	Extrapolation logarithmique	0.65	4.28
2002	Extrapolation logarithmique	-0.29	1.52
2003	Extrapolation logarithmique	-0.13	2.18
2004	Extrapolation logarithmique	-0.70	1.37

La série optimale de mesures ainsi déterminées est représentée graphiquement (en gras) et confrontée aux autres mesures déjà effectuées (Figure 51).

Figure 52 : évolution de l'irrégularité (avec courbe de tendance) comparée aux autres mesures





Il apparaît graphiquement que la mesure de référence modifie significativement les résultats constatés jusqu'alors. En effet, la modification de la méthode de mesure en

fonction de l'endroit où se situe le seuil entraîne des bouleversements. L'évolution des irrégularités devient plus difficile à décrire puisqu'elle connaît, aussi bien en ce qui concerne l'irrégularité à gauche que l'irrégularité à droite, une première partie décroissante. Puis, l'année où l'on change de méthode de calcul, les irrégularités reprennent une valeur élevée. Ce renversement de tendance est illustré par la régression polynomiale tracée sur les graphes de la Figure 52

Ces résultats peuvent sembler à première vue difficiles à interpréter. Ils permettent cependant de mettre en lumière trois résultats :

(1) L'irrégularité est située sur une portion de la courbe relativement proche du sommet. Cette proximité rend difficile l'estimation par interpolation les années où elle est trop importante.

(2) Modifier la méthode d'évaluation pour étudier une série d'irrégularités dans le temps (ici interpolation logarithmique et extrapolation logarithmique) rend difficile l'interprétation des résultats. En effet, il apparaît clairement que la tendance est rompue. L'année 2001 correspond au changement méthodologique.

(3) Enfin, le résultat le plus important, est que la réflexion sur l'allure de la distribution menée jusqu'alors est insuffisante. L'analyse repose sur le présupposé que la distribution doit avoir une allure (plus ou moins) en forme de cloche (Gaussienne). Ce présupposé doit être remis en question. C'est le point de départ de la proposition méthodologique de mesure paramétrique menée dans la partie 3 de la thèse.

e. Conclusion et apports sur les mesures d'irrégularités

Les mesures effectuées dans cette section n'ont pas pour objet l'étude de l'évolution des irrégularités en France. Cette étude est menée dans le chapitre IX. L'étude s'intéresse ici essentiellement aux conclusions méthodologiques concernant l'impact du choix de l'instrument de mesure sur l'irrégularité. Elle conduit à formuler les constats suivants :

(1) Toutes les mesures fournissent des résultats globalement cohérents. Il est important de le souligner. Il y a des différences notables année par année, mais les tendances sur une durée de 13 ans sont similaires.

(2) Il apparaît que certaines mesures sont plus fragiles car très sensibles au choix de la largeur des intervalles ou aux aléas de la distribution, surtout si la population étudiée est de taille modeste, dans la mesure où elles ne lissent pas les calculs sur un grand nombre

d'intervalles. Il s'agit des deux mesures par symétrie, et de la mesure par moyenne des deux intervalles adjacents. La fragilité de ces mesures est importante à souligner dans la mesure où ce sont pourtant les mesures les plus souvent utilisées dans la littérature comptable.

(3) À l'inverse, les mesures par interpolation, si elles fournissent des résultats plus satisfaisants, soulèvent trois séries de problèmes. (a) Si le lissage permet d'éviter la volatilité des mesures par symétrie ou par moyenne des deux intervalles adjacents, une étendue de référence trop large risque de créer une difficulté de représenter correctement l'endroit de la courbe où l'on se situe en « aplatissant » toutes les données. (b) Ces mesures ne sont pas adaptées lorsque l'on se situe à proximité du sommet de la courbe, et les mesures par extrapolation qui théoriquement devraient prendre le relais sont délicates à mettre en œuvre dans la mesure où demeure une incertitude sur l'allure de la courbe. (c) Enfin, le choix entre les trois méthodes d'interpolation (comme avec les méthodes d'extrapolation) est délicat car il demande de faire une hypothèse sous jacente sur l'allure théorique de la distribution. La mesure par interpolation linéaire a tendance à être une mesure plus moyenne, sous-estimant les deux irrégularités, alors que l'interpolation exponentielle a tendance à « creuser » l'allure de la courbe de référence, et donc à surestimer l'irrégularité à droite et sous-estimer l'irrégularité à gauche. L'interpolation logarithmique a l'effet inverse.

En conclusion, les résultats permettent de répondre aux deux questions initialement formulées. (1) Ils confirment que le choix d'une méthode influence de manière importante les résultats des mesures, et par là même les conclusions d'une étude comparant les seuils. Il semble dès lors possible de revisiter les résultats obtenus jusqu'alors par les auteurs en comptabilité qui se sont intéressés au sujet. (2) Par ailleurs, ils permettent de porter un jugement sur les différentes méthodes recensées dans la littérature. Les méthodes par symétrie ou par moyenne arithmétique simple, pourtant les plus utilisées dans la littérature, apparaissent comme étant les plus fragiles. Les mesures par interpolation apparaissent plus appropriées, et en l'absence de postulat sur l'allure attendue de la distribution, l'interpolation linéaire, qui fournit des mesures intermédiaires entre interpolation logarithmique et exponentielle, est sans doute la mesure la plus prudente.

Enfin, seule une réflexion sur l'allure de la courbe de distribution peut conduire à améliorer les mesures. Cette réflexion est menée dans la partie 3 de la thèse.

2. Étude des indicateurs d'asymétrie

Après avoir examiné les mesures d'irrégularité fondées sur l'estimation d'un effectif théorique, les indicateurs d'asymétrie, qui se contentent de comparer des effectifs observés entre eux, sont examinés. Pour ce faire, le degré d'indépendance de l'irrégularité à gauche et à droite du seuil de zéro est mesuré. En effet, les auteurs utilisant pour comparer les irrégularités des indicateurs confondant les deux phénomènes, ces indicateurs ne sont théoriquement pertinents que si les deux mesures sont fortement corrélées.

a. Comparaison de l'évolution de l'irrégularité à droite et de l'irrégularité à gauche du seuil résultat nul

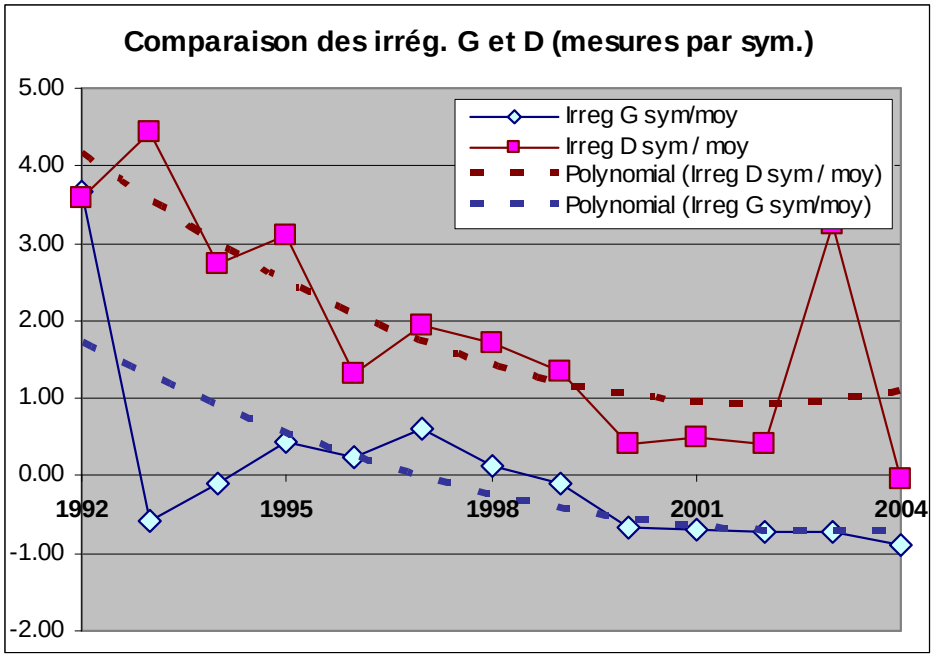
Les mesures des irrégularités à droite et à gauche du seuil sont comparées sur la base de deux méthodes[90] : la symétrie par rapport à la moyenne et l'interpolation linéaire.

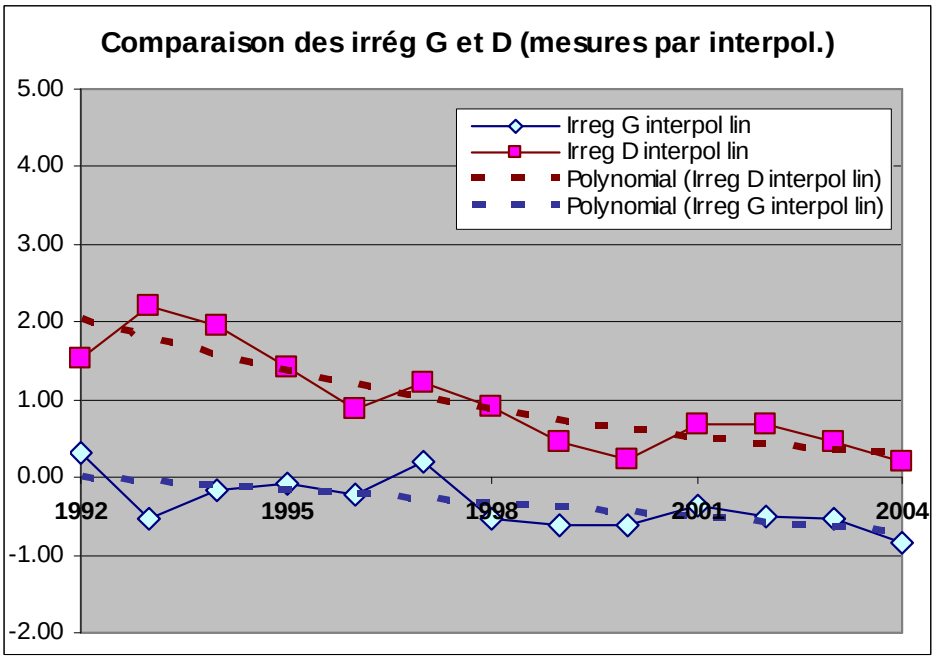
Tableau 35 : Mesures de l'irrégularité (relative) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire

Année	Irrégularité relative à gauche		Irrégularité relative à droite	
	mesurée par symétrie par rapport à la moyenne	mesurée par interpolation linéaire	mesurée par symétrie par rapport à la moyenne	mesurée par interpolation linéaire
1992	367 %	32 %	357 %	154 %
1993	-58 %	-53 %	443 %	219 %
1994	-9 %	-17 %	273 %	194 %
1995	44 %	-7 %	310 %	142 %
1996	25 %	-23 %	132 %	88 %
1997	62 %	20 %	194 %	121 %
1998	13 %	-54 %	171 %	91 %
1999	-11 %	-61 %	133 %	44 %
2000	-65 %	-62 %	40 %	23 %
2001	-70 %	-37 %	50 %	68 %
2002	-73 %	-49 %	40 %	67 %
2003	-73 %	-55 %	325 %	45 %
2004	-90 %	-83 %	-4 %	19 %

[90] Pour alléger les calculs, les 11 mesures étudiées dans la section précédente n'ont pas été utilisées. Seules ont été retenues la symétrie et l'interpolation linéaire (qui est une extension de la moyenne arithmétique) car ces deux méthodes sont le plus souvent utilisées dans la littérature comptable et l'objet de cette partie de l'étude est de montrer les limites de ces méthodes.

Figure 53 : Comparaison graphique de l'évolution des irrégularités (en valeur absolue) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire



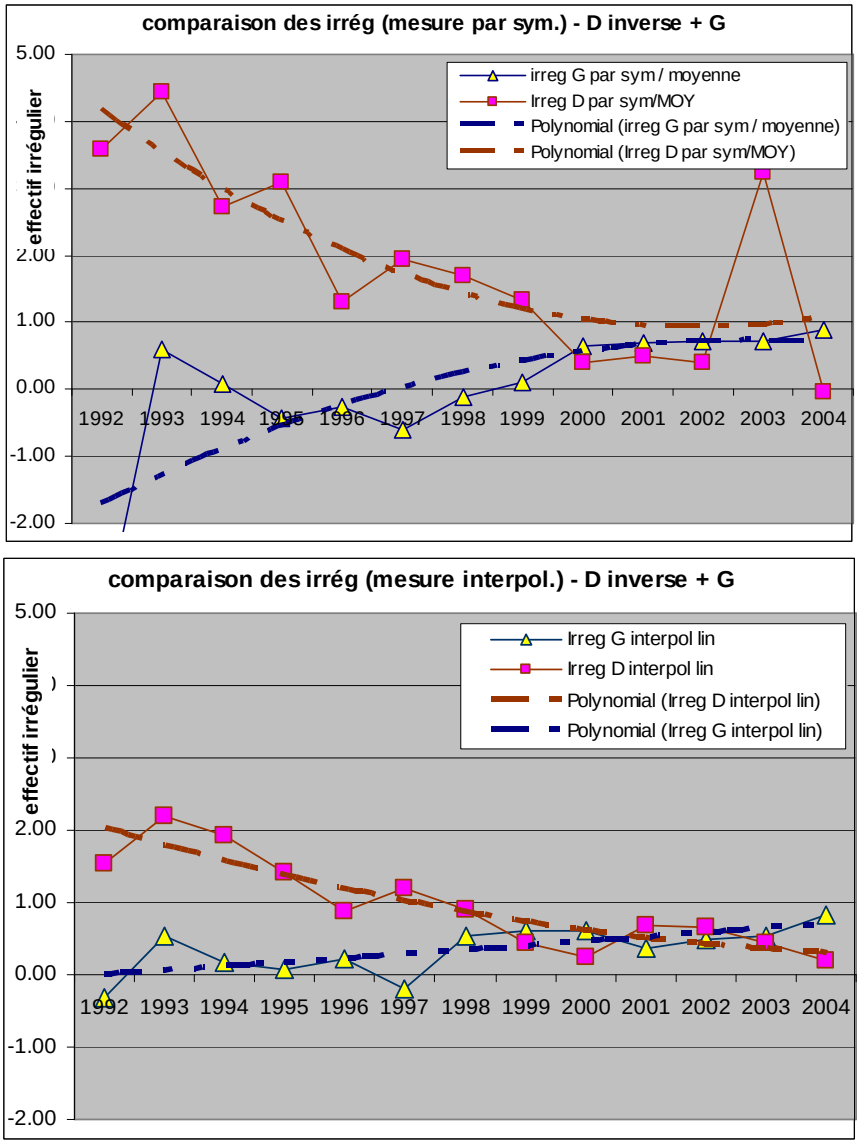


Dans les deux cas, une observation graphique sommaire des irrégularités droite et gauche peut laisser penser que leurs évolutions sont similaires : une baisse des irrégularités, plus importante en début de période, et une certaine stabilisation après l'année 2000[91].

[91] La section précédente a montré que toutes les mesures aboutissent au même constat.

Mais cette comparaison graphique ne tient pas compte du signe de l'irrégularité : en effet, l'irrégularité positive à droite signifie qu'il y a un sur-effectif, alors que l'irrégularité à gauche, généralement négative, s'interprète comme un sous-effectif. Autrement dit, une diminution de la valeur négative de l'irrégularité signifie qu'il y a accroissement de l'irrégularité en valeur absolue. Pour tenir compte du signe de l'irrégularité et représenter les deux irrégularités sur le même graphe, l'irrégularité à gauche du seuil est représentée en inversant son signe.

Figure 54 : Comparaison graphique de l'évolution des irrégularités (tenant compte du signe de l'irrégularité) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire



Cette nouvelle représentation révèle de manière flagrante que l'évolution de l'irrégularité à gauche, dans l'échantillon d'entreprises françaises entre 1992 et 2004, évolue de manière

strictement opposée à l'irrégularité à droite du seuil. Autrement dit, alors que l'irrégularité à droite diminue (le sur-effectif est plus important en début de période), l'irrégularité à gauche augmente (la sous-représentation des entreprises faiblement déficitaires, qui était pratiquement nulle en début de période, devient de plus en plus marquée durant la période d'étude).

En conclusion, l'étude montre que l'irrégularité à droite n'est pas corrélée positivement à l'irrégularité à gauche du seuil[92]. Cela signifie que l'étude de chacune de ces irrégularités mérite d'être menée indépendamment de l'autre. Ce résultat conduit à remettre en cause les fondements sur lesquels reposent les calculs d'asymétrie. Ces calculs sont effectués dans le paragraphe suivant.

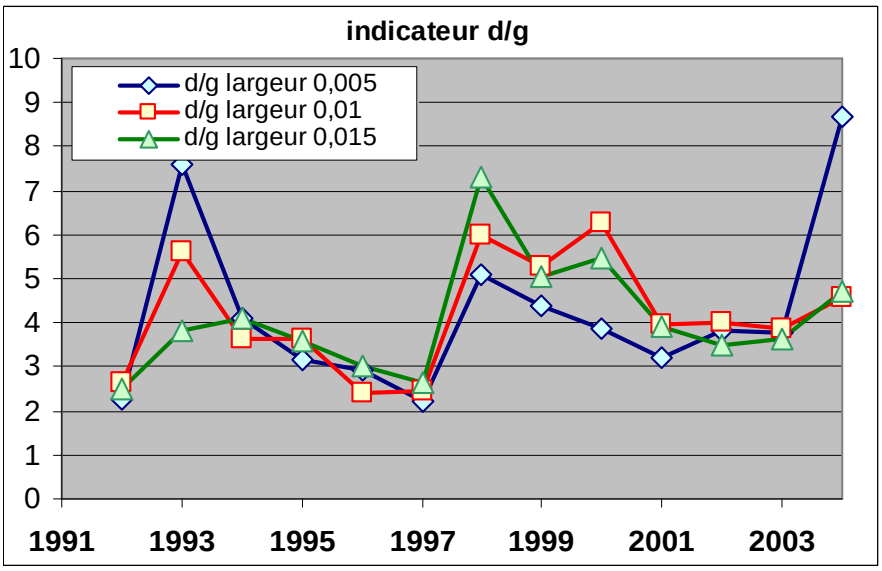
b. Mesure des indicateurs d'asymétrie et comparaison

Les auteurs qui utilisent un indicateur d'asymétrie comme outil de mesure des irrégularités et de comparaison dans le temps ou dans l'espace (Dechow, Richardson et Tuna, 2003 ; Leuz, Nanda et Wysocki, 2003 ; Glaum, Lichtblau et Lindemann, 2004 ; Burgstahler, Hail et Leuz, 2006) ne distinguent pas les deux irrégularités à droite et à gauche du seuil. Leurs mesures sont appliquées à l'échantillon d'entreprises françaises de 1992 à 2004.

Dans un premier temps, l'indicateur d'asymétrie d/g (Dechow, Richardson et Tuna, 2003 ; Leuz, Nanda et Wysocki, 2003 ; Burgstahler, Hail et Leuz, 2006) est mesuré. Dans le souci de contrôler les résultats, trois calculs sont effectués en faisant varier la largeur des intervalles.

[92] Si corrélation il y a, elle est négative.

Figure 55 : évolution de l'indicateur d'asymétrie d/g entre 1992 et 2004



Logiquement, il apparaît que plus la largeur est élevée, moins l'indicateur est volatile, ce qui est normal dans la mesure où plus les effectifs sont réduits, plus les variations d'effectifs ont un impact important. La tendance reste cependant la même.

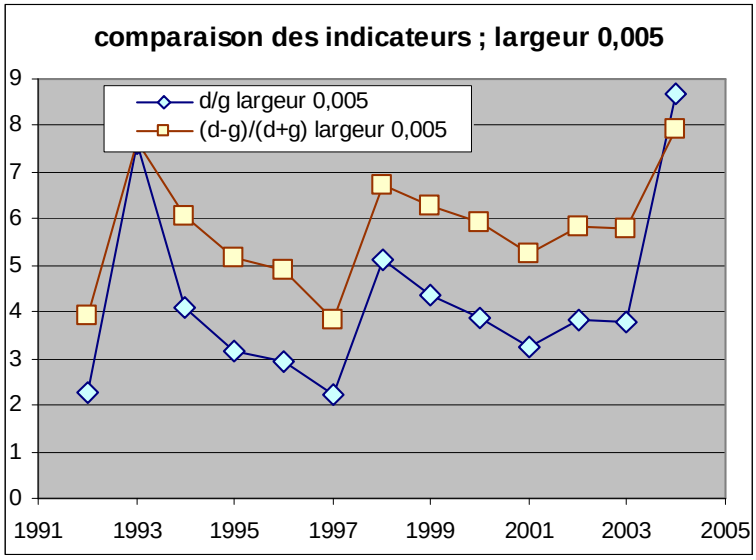
Cet indicateur est ensuite comparé avec l'indicateur (d-g)/(d+g) (Glaum, Lichtblau et Lindemann, 2004). Comme l'indicateur d/g est compris entre 0 (si l'effectif de droite est nul) et + ∞, alors que l'indicateur (d-g)/(d+g) est compris entre -1 et 1, le second indicateur est multiplié par 10 pour permettre une comparaison graphique (voir Tableau 36).

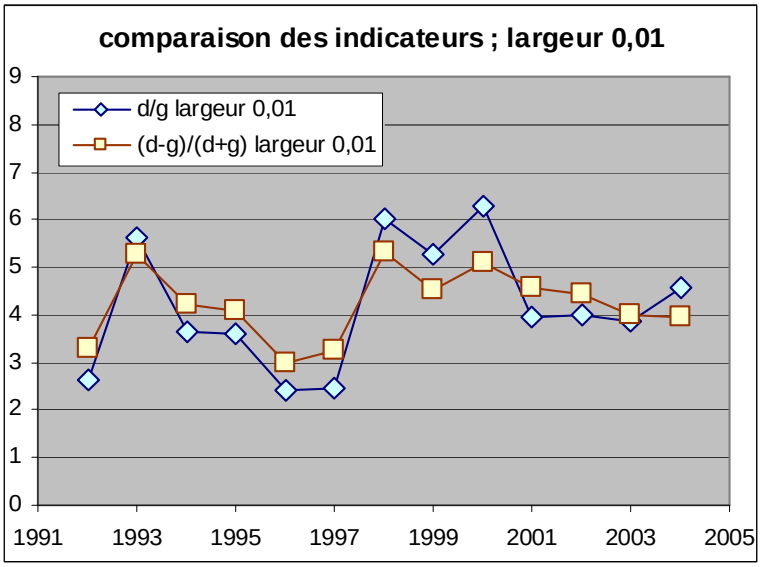
Tableau 36 : Calcul des indicateurs d'asymétrie d/g et (d-g)/(d+g) ; entreprises françaises cotées de 1992 à 2004

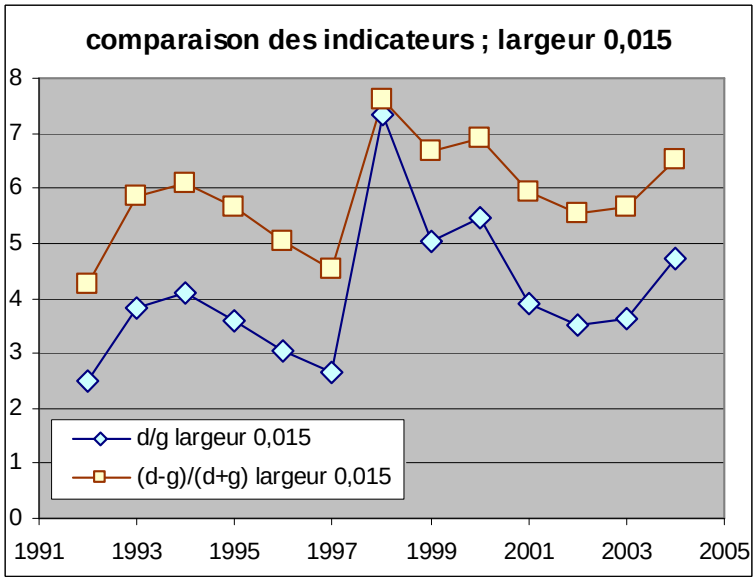
	effectifs						d/g			10 x (d-g)/(d+g)		
	[-0,015 ;-0,00[	[-0,01 ;-0,00[	[-0,005 ;0[	[0 ; 0,005[	[0 ; 0,01[	[0 ; 0,015[	largeur 0,005	largeur 0,01	largeur 0,015	largeur 0,005	largeur 0,01	largeur 0,015
1992	27	19	14	32	50	67	2.29	2.63	2.48	3.91	3.30	4.26
1993	18	10	5	38	56	69	7.60	5.60	3.83	7.67	5.29	5.86
1994	21	17	10	41	62	86	4.10	3.65	4.10	6.08	4.21	6.07
1995	25	18	13	41	65	90	3.15	3.61	3.60	5.19	4.09	5.65
1996	38	33	15	44	79	115	2.93	2.39	3.03	4.92	3.01	5.03
1997	41	34	21	47	83	109	2.24	2.44	2.66	3.82	3.27	4.53
1998	18	16	9	46	96	132	5.11	6.00	7.33	6.73	5.33	7.60
1999	22	14	8	35	74	111	4.38	5.29	5.05	6.28	4.51	6.69
2000	24	15	9	35	94	131	3.89	6.27	5.46	5.91	5.10	6.90
2001	29	22	13	42	87	113	3.23	3.95	3.90	5.27	4.58	5.92
2002	36	24	11	42	96	126	3.82	4.00	3.50	5.85	4.44	5.56
2003	31	20	9	34	77	112	3.78	3.85	3.61	5.81	3.99	5.66
2004	19	12	3	26	55	90	8.67	4.58	4.74	7.93	3.94	6.51

Comme le montre la Figure 56 les deux indicateurs fournissent une information similaire. Les limites de l'indicateur d/g (Glaum, Lichtblau et Lindemann, 2004) ne se font sentir que sur des valeurs extrêmes (effectif nul de l'effectif de gauche notamment) qui ne sont pas apparues dans le cadre de l'application. Par ailleurs, l'indicateur (d-g)/(d+g) n'étant jamais négatif, la représentation sur un même graphe des deux indicateurs est possible.

Figure 56 : Comparaison graphique des indicateurs d'asymétrie d/g et (d-g)/(d+g) en faisant varier la largeur des intervalles

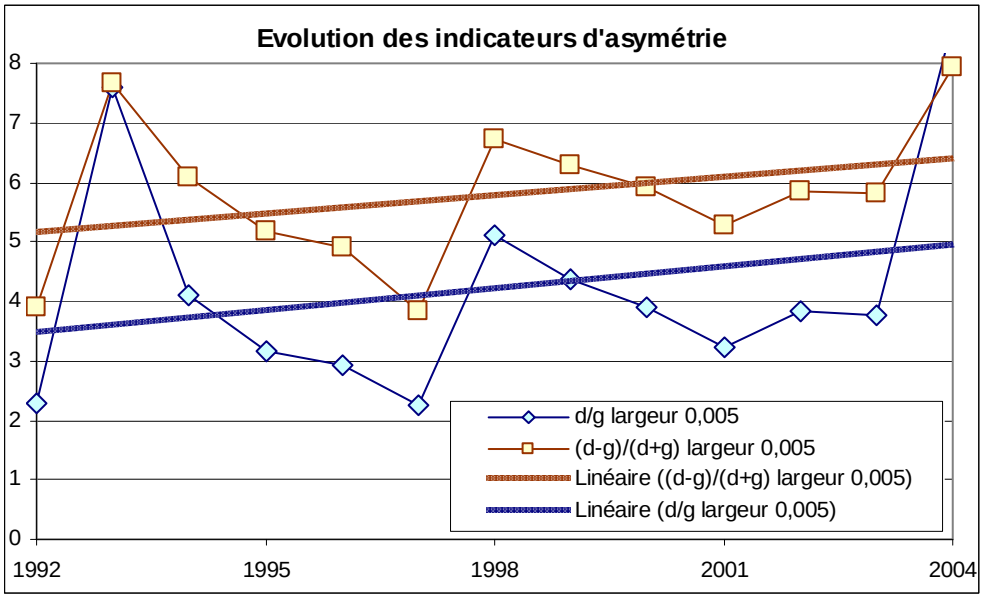






Enfin, comme l'ont fait Dechow, Richardson et Tuna (2003), l'évolution de l'irrégularité mesurée par asymétrie est observée. La Figure 57 fait apparaître la droite de régression permettant de visualiser la tendance de chaque série correspondant à chacun des indicateurs d'asymétrie.

Figure 57 : Comparaison graphique des indicateurs d'asymétrie d/g et (d-g)/(d+g) avec courbe de tendance



Il apparaît nettement que l'indicateur d'asymétrie, quelle que soit la méthode utilisée, augmente durant la période étudiée. Selon la littérature comptable, cela doit être interprété comme la manifestation d'une aggravation des pratiques de gestion du résultat pour éviter les seuils. Or ce résultat masque le double phénomène déjà mis en évidence (dans la sous-partie a) : une diminution de l'irrégularité à droite et l'augmentation de l'irrégularité à gauche.

En conclusion, il apparaît que les indicateurs d'asymétrie sont myopes. La diminution de l'irrégularité à droite du seuil (qui se traduit par une diminution des effectifs) est masquée par l'augmentation de l'irrégularité à gauche (qui se traduit également par une diminution des effectifs). Ce biais méthodologique est d'autant plus fort que les effectifs à gauche étant en général plus faibles, une diminution de même ampleur des effectifs à droite et à gauche induisent une augmentation non proportionnelle de l'asymétrie puisque le dénominateur est formé d'un chiffre plus faible. L'utilisation des indicateurs d'asymétrie ne semble donc pas satisfaisante dans l'étude des pratiques de gestion du résultat pour éviter les seuils.

c. Une relecture des travaux antérieurs

Dans leur étude sur l'évolution des seuils pour les entreprises cotées au NYSE de 1989 à 2001, Dechow, Richardson et Tuna (2003) observent que l'indicateur d'asymétrie au seuil résultat nul décline régulièrement de 2,09 à 1,31 lorsqu'en 1995 les conditions de cotation[93] changent. Mais les auteurs ne fournissent pas de détail de l'impact sur les irrégularités respectivement à gauche et à droite du seuil. Par ailleurs, Leuz, Nanda et Wysocki (2003) comme Burgstahler, Hail et Leuz (2006) étudient la gestion du résultat dans 31 pays et les classent notamment grâce à un indicateur d'asymétrie. Tous ces résultats peuvent être revisités.

Par ailleurs, l'observation graphique (voir la Figure 13 de la page 74) des résultats de Glaum, Lichtblau et Lindemann (2004) laisse penser que la non significativité des tests de comparaison des entreprises allemandes et américaines fondées sur un indicateur d'asymétrie masque en réalité une double divergence. Les entreprises allemandes légèrement bénéficiaires sont sur-représentées, alors que les entreprises américaines légèrement déficitaires sont sous-représentées. L'indicateur composite annule, par compensation, les deux différences. La mise en lumière de cette double irrégularité ouvre donc de nouvelles perspectives de recherche.

d. Conclusion et apports sur les indicateurs d'asymétrie

Seule une forte corrélation entre les deux irrégularités à droite et à gauche du seuil permet de légitimer l'utilisation d'indicateurs d'asymétrie. Cette corrélation n'est cependant pas vérifiée sur l'échantillon d'entreprises cotées françaises sur la période 1992 à 2004. Il semble donc que l'utilisation d'indicateurs d'asymétrie pour mesurer la gestion du résultat pour éviter les seuils doit être évitée. Ce résultat a un intérêt évident pour les chercheurs qui s'engagent dans des études comparatives, internationales ou longitudinales.

Plus fondamentalement, il semble que le principal problème posé par l'utilisation d'indicateurs d'asymétrie est un problème conceptuel. Cet indicateur considère en effet que l'effet de seuil se définit par la somme des deux effets : sur-représentation à droite, et sous-représentation à gauche. L'étude montre qu'il est utile de ne pas confondre ces deux

[93] « Listing requirements » dans l'article en anglais.

phénomènes qui peuvent avoir des évolutions distinctes. Une porte est ouverte à de futures études sur ce sujet.

Enfin, d'autres limites doivent être soulignées. L'indicateur d'asymétrie ne s'affranchit pas des problèmes de largeur d'intervalle. Ce problème est d'autant plus important que l'indicateur a été utilisé notamment pour comparer des populations de taille différente, et pour lesquelles la largeur d'intervalle retenue peut être différente (Leuz, Nanda et Wysocki, 2003 ; Burgstahler, Hail et Leuz, 2006). Par ailleurs, les mesures sont faites sur la totalité des effectifs d'un intervalle, et non pas sur les effectifs irréguliers, ce qui peut diluer la mesure et provoquer des problèmes d'échelle. La mesure de l'asymétrie ne tient pas compte de l'endroit où l'on se trouve sur la courbe, et ne règle pas le problème de la pente non linéaire de la courbe de distribution, rendant toute interprétation délicate. Enfin, les indicateurs d'asymétrie peuvent être, par construction, totalement inopérants dans des cas extrêmes, notamment lorsque le dénominateur est nul.

B. Impact de la largeur des intervalles, et nouvel indicateur d'irrégularité

L'objectif de cette section est de montrer que le choix de la largeur de l'intervalle[94] d'observation a un impact important sur la mesure de l'irrégularité. Les irrégularités à droite et à gauche du seuil du résultat nul sont mesurées en faisant varier la largeur d'intervalle et selon les deux méthodes utilisées par Burgstahler et Dichev (1997), à savoir la moyenne des deux intervalles adjacents, et la symétrie par rapport à la moyenne. Les calculs ne sont effectués que pour une seule année pour alléger les calculs[95].

1. Choix d'une année d'étude et observation graphique

a. Choix de l'année 1997

Dans le cadre de cette section, l'année d'étude a été sélectionnée parmi les 13 années de la base en tenant compte des contraintes techniques liées aux calculs par symétrie. Pour effectuer les mesures par symétrie, il est nécessaire de déterminer un axe de symétrie, correspondant au sommet de la distribution. Or plusieurs choix sont envisageables : la moyenne, le mode ou la médiane.

Le mode est à proscrire dès lors que les irrégularités (notamment au seuil de zéro) peuvent conduire la distribution à être multimodale. Par ailleurs, moyenne et médiane présentent selon les années des différences sensibles. Pour contourner la difficulté, dans le cadre de l'étude sur les largeurs des intervalles, l'année retenue est celle où médiane et moyenne sont les plus proches, soit 1997.

[94] En français, on peut trouver l'expression « largeur des classes d'observation », ou « largeur des pas ». En anglais, on trouve les expressions « interval width », « bin width » ou « classes range ».

[95] Les mêmes mesures auraient pu être effectuées pour chaque année, et en multipliant les instruments de mesure. Mais ce travail chronophage eut été inutile puisque l'objet de l'étude est ici de montrer qu'il existe un impact. Si l'impact est révélé sur une année au moins, la conclusion est validée.

Tableau 37 : Rappel des paramètres des distributions annuelles de IB/(AT-IB)

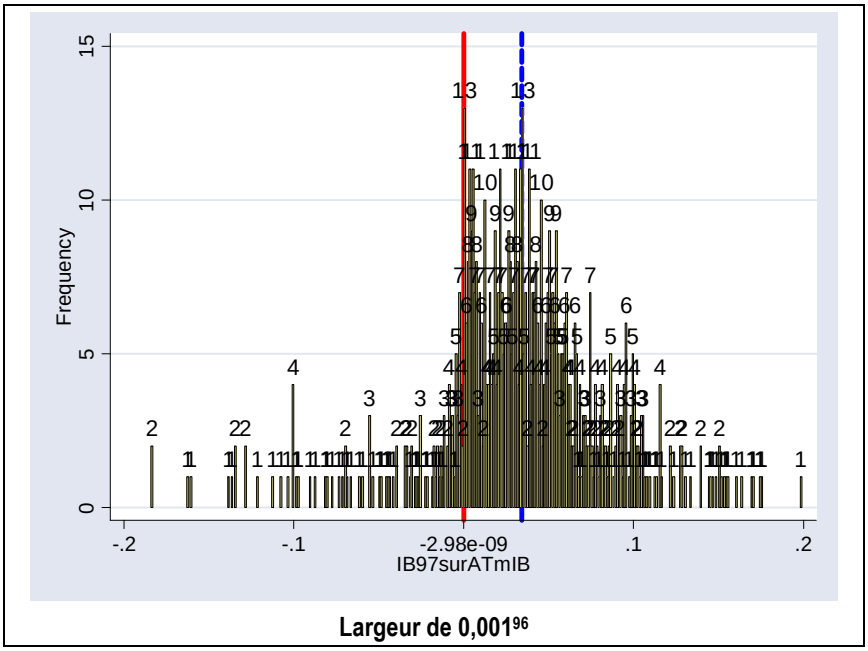
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Nombre	291	296	376	422	616	702	753	798	791	734	698	659	610
Moyenne	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0463	0.0233	0.0014	0.0012	-0.0002	0.0178
Médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.0350	0.0346	0.0281	0.0214	0.0172	0.0185	0.0259
écart entre moyenne et médiane	-0.0083	-0.0026	-0.0041	-0.0026	-0.0036	**-0.0008**	-0.0056	-0.0117	0.0048	0.02	0.016	0.0187	0.0081

L'année 1997 présente également d'autres avantages : la moyenne n'est pas confondue avec le résultat nul d'une part et le nombre d'observations est relativement élevé (702, le maximum étant en 1999 avec 798 observations).

b. Observations graphiques de la distribution en fonction de la largeur des intervalles

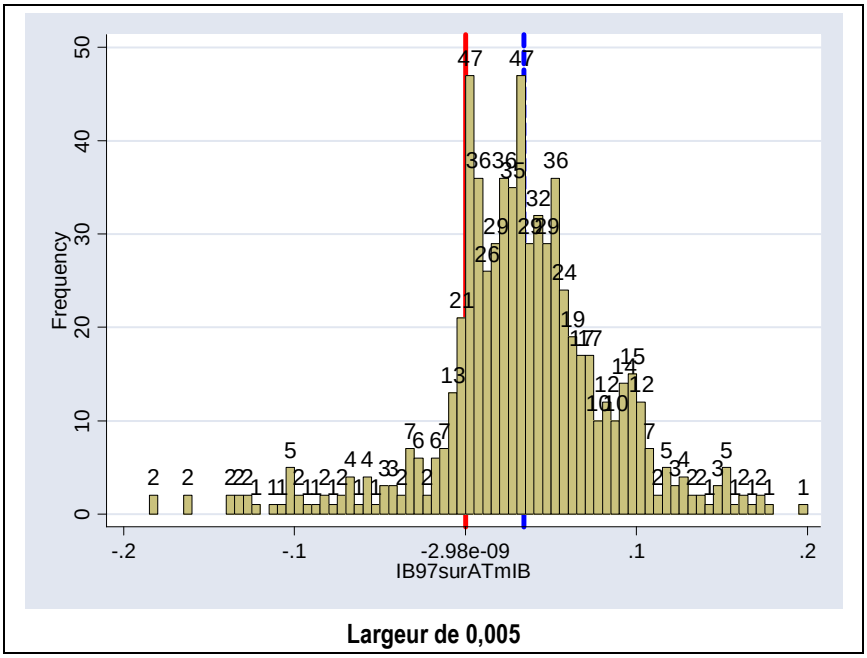
Les deux graphes de la Figure 58 représentent la distribution totale avec une largeur d'intervalle très faible (respectivement 0,001 et 0,005). La ligne en gras (rouge) figure le seuil du résultat nul. La ligne en pointillé et gras (bleue) figure la moyenne.

Figure 58 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles faibles ; 0,001 et 0,005)



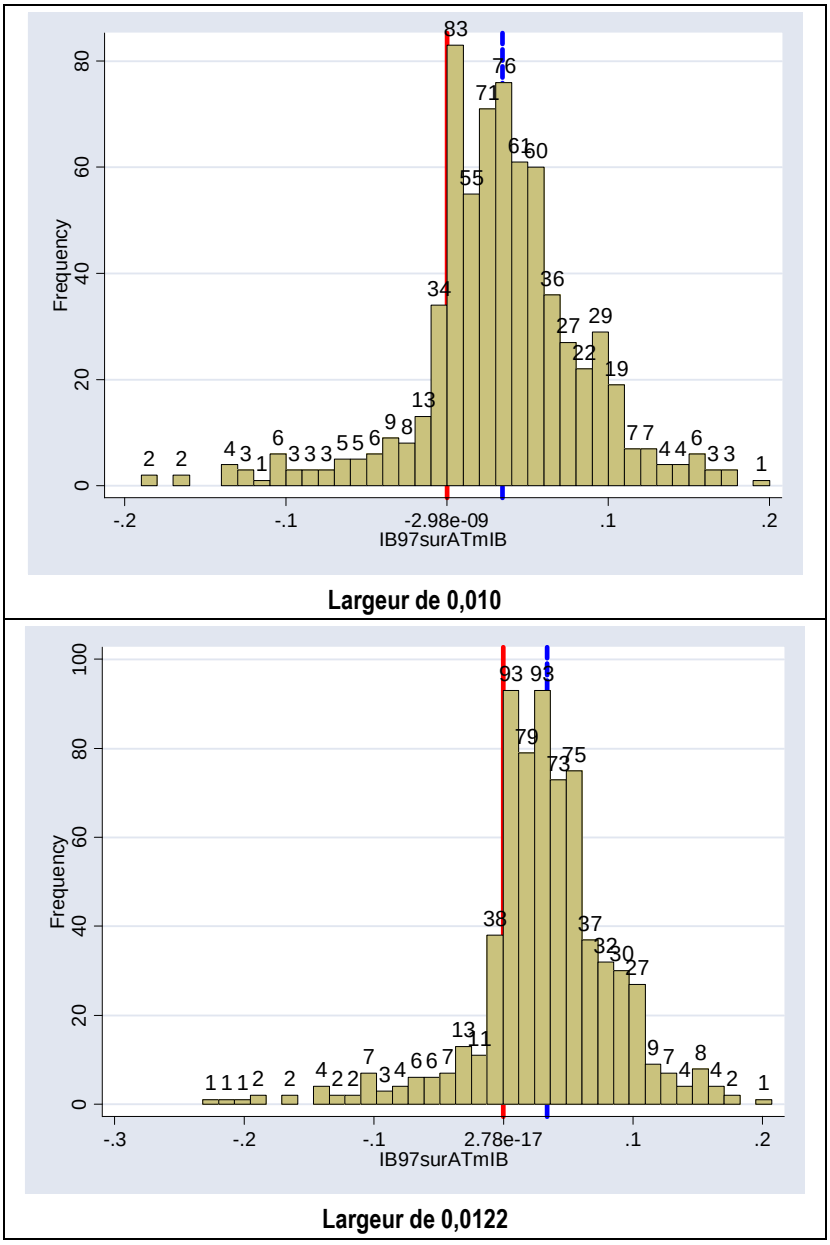
Largeur de 0,001[96]

[96] Fonction sous Stata : histogram IB97surATmIB if IB97surATmIB>-0.2 & IB97surATmIB<0.2, width(0.001) start(-0.20000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lwidth(thick) lcolor(red)) xline(0.0341, lwidth(thick) lcolor(blue) lpattern(dash))



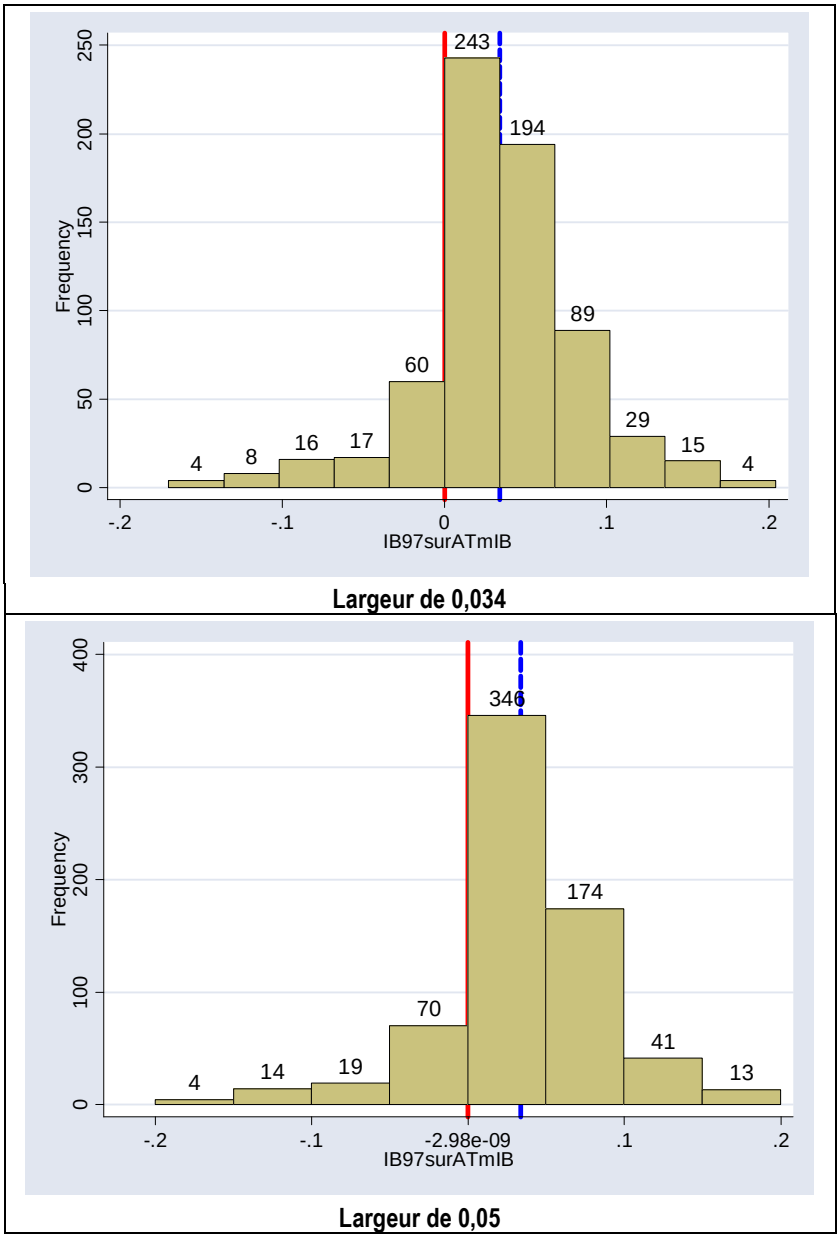
Il apparaît graphiquement que la distribution est d'autant moins lisse que la largeur d'observation est réduite. Cela va influencer l'estimation de l'irrégularité par calcul de moyenne ou par symétrie. Il apparaît également que la distribution est multimodale. Cela apparaît nettement lorsque la largeur des intervalles est comprise entre 0,005 et 0,01 (voir Figure 59). Le seuil étudié est situé sur le principal mode de la distribution. Cela confirme l'ampleur de l'irrégularité. Il semble en revanche (graphiquement) qu'en 1997 l'irrégularité à gauche du résultat nul ne soit pas très prononcée.

Figure 59 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles intermédiaires ; 0,010 et 0,0122)



La largeur d'intervalle de 0,0122 correspond à la largeur théorique (Silverman, 1986 ; Scott, 1992). Graphiquement, il apparaît que cette largeur ne maximise pas les irrégularités.

Figure 60 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles importantes ; 0,034 et 0,050)



Sur la Figure 60, le graphe de gauche est tracé avec une largeur égale à la moyenne, à savoir 0,034. Le second graphe représente une largeur supérieure à la moyenne. Lorsque la largeur égale la moyenne, l'intervalle supérieur au seuil atteint le sommet de la distribution. Une telle largeur permet encore de distinguer l'irrégularité dans la mesure où l'intervalle situé juste au-dessus du seuil demeure le plus élevé de la distribution. En revanche, lorsque la largeur dépasse la distance entre moyenne et résultat nul, il n'est plus possible de distinguer d'irrégularité puisque celle-ci se confond avec le sommet de la courbe.

En conclusion, la simple observation graphique permet de souligner les difficultés liées à la sélection d'une largeur d'intervalles pour calculer les irrégularités de distribution :

(1) Les petites largeurs conduisent à des mesures erratiques dans la mesure où la distribution n'est pas lisse, et que le nombre d'observations par intervalle devient faible.

(2) La largeur théorique (Silverman, 1986 ; Scott, 1992) ne conduit pas nécessairement à optimiser les mesures.

(3) Élargir l'intervalle au-delà du sommet de la distribution rend les mesures inopérantes.

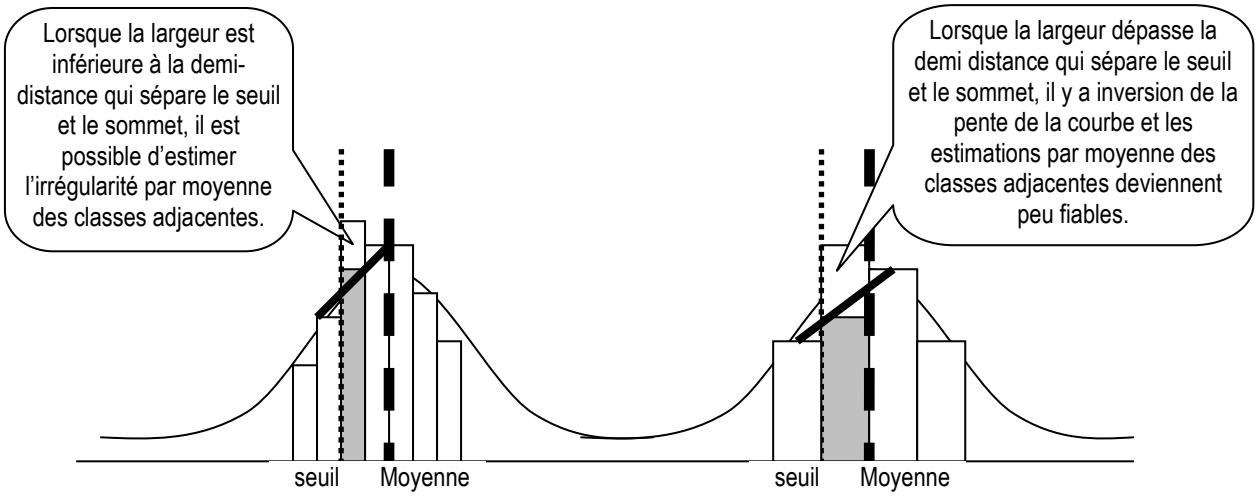
c. Limites des largeurs des intervalles

Dans cette étude, la largeur d'intervalle varie depuis un minimum de 0,001 (inférieur à la valeur de 0,005 généralement retenue dans la littérature) jusqu'à une valeur maximale qu'il faut déterminer. Il apparaît que cette valeur maximale est techniquement contrainte par le mode de calcul retenu.

c.1. Contrainte pour le calcul par moyenne

Comme les deux graphes de la Figure 60 l'ont montré, lorsque la largeur de l'irrégularité à droite du seuil atteint le sommet de la distribution, il y a inversion de la pente de la distribution. Toute estimation par moyenne conduit alors à surestimer l'irrégularité.

Figure 61 : Contrainte de largeur maximale dans le calcul par moyenne



Autrement dit, dans le calcul de l'irrégularité par moyenne, il y a une limite mécanique maximum à la largeur d'intervalle égale à la demi-distance entre le sommet de la distribution et le seuil, puisque dès cette valeur, l'intervalle adjacent supérieur à l'intervalle estimé empiète sur la partie de la distribution pour laquelle il y a inversion de la pente. Quand le seuil étudié est celui du résultat nul, la largeur limite est la valeur absolue de la demi-moyenne (ou de la demi-médiane selon l'indicateur retenu comme représentant le sommet de la distribution[97]).

Tableau 38 : Largeur maximale année par année dans le cas de l'étude des entreprises françaises cotées ; calcul par moyenne arithmétique

	1992	1993	1994	1995	**1996**	**1997**	**1998**	**1999**	2000	**2001**	**2002**	**2003**	2004
Moyenne/2	0.0140	0.0099	0.0145	0.0126	0.0160	0.0170	0.0203	0.0231	0.0117	**0.0007**	**0.0006**	**-0.0001**	0.0089
Médiane/2	0.0099	0.0087	0.0125	0.0114	0.0143	0.0167	0.0175	0.0173	0.0141	0.0107	0.0086	0.0093	0.0129
Largeur théorique ($2IQRn^{-1/3}$)	0.0143	0.0155	0.0138	0.0137	**0.0128 ***	**0.0122 ***	**0.0139 ***	**0.0136 ***	0.0120	0.0132	0.0144	0.0141	0.0128
	* : une étoile signifie que la largeur théorique peut être utilisée pour calculer l'irrégularité par moyenne arithmétique des deux classes adjacentes												

Il est important de souligner que retenir la largeur théorique d'intervalle conduit presque chaque année à interdire l'estimation de l'irrégularité par moyenne arithmétique des deux classes adjacentes (la largeur théorique est presque toujours supérieure aux maximums techniquement autorisés). De plus, si la moyenne est retenue comme sommet de la distribution, alors il apparaît que les années où la moyenne est proche de zéro (essentiellement entre 2001 et 2003), les calculs par moyenne arithmétique sont totalement inadaptés même pour la largeur réduite de 0,005 généralement retenue dans la littérature.

Enfin, le même type de contrainte pèse sur les calculs par interpolation lorsque les intervalles de référence empiètent sur l'axe de symétrie.

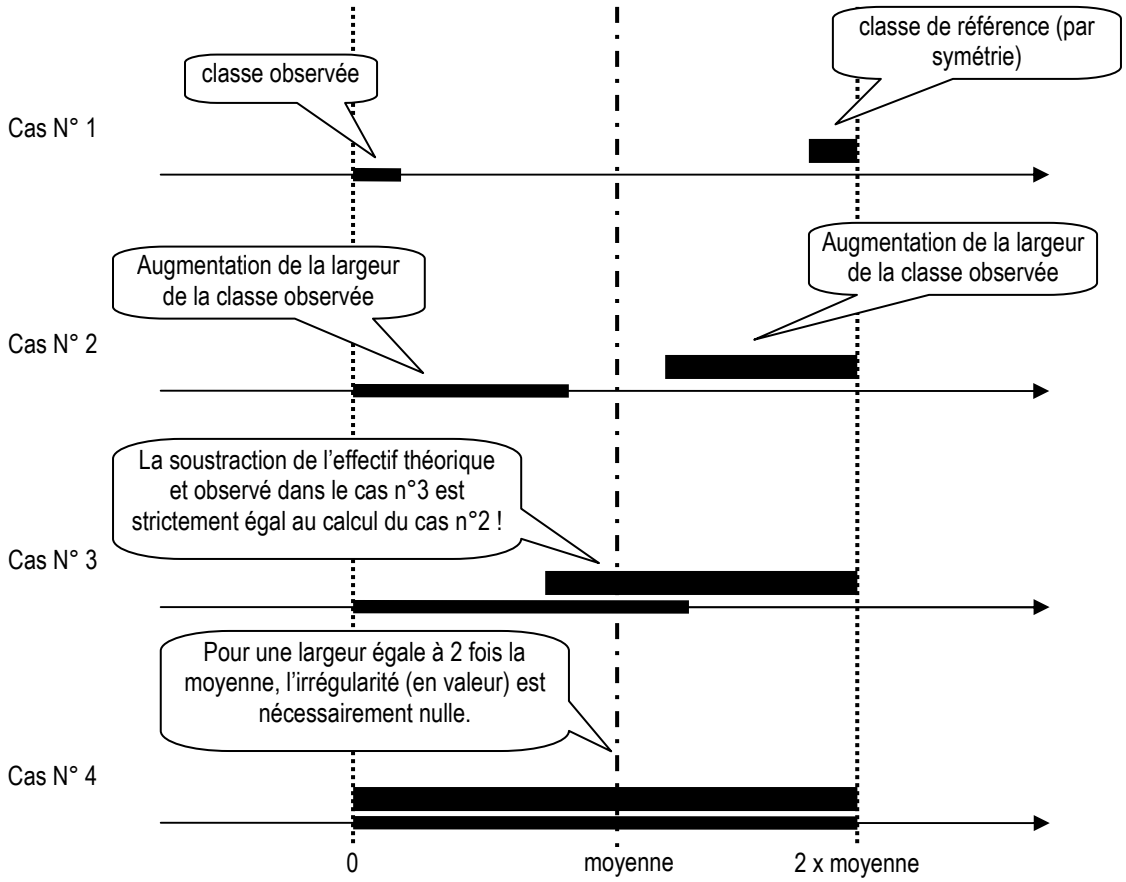
Dans le cadre de l'étude de l'année 1997, les mesures par moyenne des deux classes adjacentes peuvent varier de 0,001 jusqu'à 0,017.

[97] Dans le cadre de cette section, le problème est écarté en retenant l'année où la différence entre moyenne et médiane était quasi nulle. Mais ce problème est étudié dans la partie 3.

c.2. Contrainte pour le calcul par symétrie

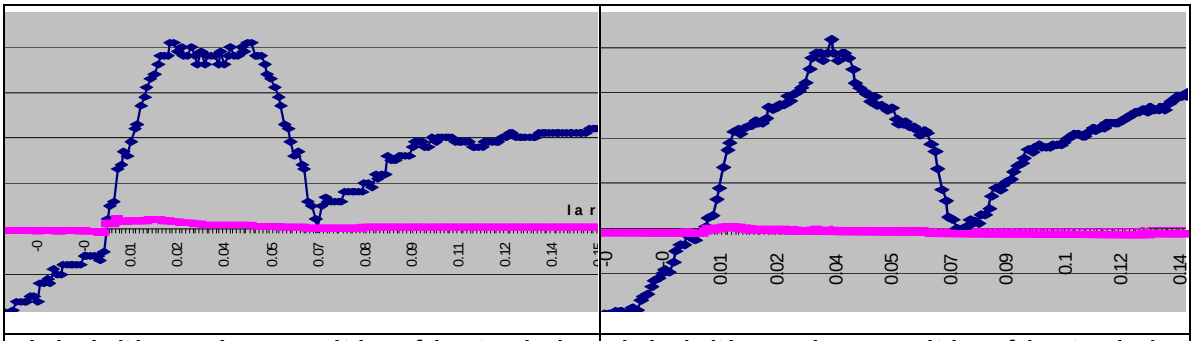
Par construction, les mesures par symétrie ont un défaut : pour mesurer l'effectif théorique de l'intervalle supérieure à zéro, et en augmentant la largeur de l'intervalle, il arrive un moment où l'intervalle de référence (théorique) devient égal à l'intervalle observé. C'est ce qu'illustre le cas 4 de la Figure 62.

Figure 62 : Contrainte de largeur maximale dans le calcul par symétrie



Dès que la largeur dépasse la distance entre le seuil et l'axe de symétrie (cas 3 ci-dessus), la différence entre effectif théorique et effectif observé devient aberrante car les deux effectifs se chevauchent. Autrement dit, la mesure par symétrie est inadaptée pour une largeur d'intervalle dépassant la distance entre le seuil étudié et l'axe de symétrie.

Figure 63 : Mesure par symétrie de l'irrégularité de distribution autour du résultat nul ; entreprises françaises cotées du SBF 250 en 1994 et en 1998



Irrégularité mesurée par symétrie en faisant varier la largeur de l'intervalle ; entreprises du SBF 250 ; année 1994	Irrégularité mesurée par symétrie en faisant varier la largeur de l'intervalle ; entreprises du SBF 250 ; année 1998

Deux illustrations de ce phénomène sont révélées lors de l'étude exploratoire menée en 2005 auprès des entreprises du SBF 250. La Figure 63 illustre ce problème. Pour une valeur comprise entre la moyenne et 2 fois la moyenne, l'irrégularité est rigoureusement symétrique à celle mesurée entre l'intervalle minimum et une valeur égale à la moyenne.

Tableau 39 : Largeur maximale année par année dans le cas de l'étude des entreprises françaises cotées ; calcul par symétrie

	1992	1993	1994	1995	1996	1997	1998	1999	2000	**2001**	**2002**	**2003**	2004
Moyenne	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0463	0.0233	**0.0014**	**0.0012**	**-0.0002**	0.0178
médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.0350	0.0346	0.0281	0.0214	0.0172	0.0185	0.0259
Largeur théorique ($2IQRn^{-1/3}$)	**0.0143 ***	**0.0155 ***	**0.0138 ***	**0.0137 ***	**0.0128 ***	**0.0122 ***	**0.0139 ***	**0.0136 ***	**0.0120 ***	0.0132	0.0144	0.0141	**0.0128 ***

* : une étoile signifie que la largeur théorique peut être utilisée pour calculer l'irrégularité par symétrie

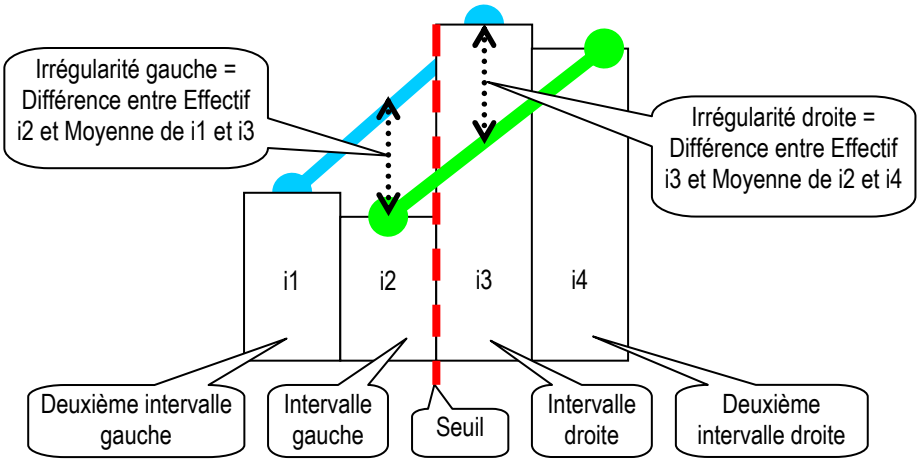
La lecture du Tableau 39 indique que, comme dans le cas des mesures par moyenne entre deux intervalles adjacents, mesurer l'irrégularité par symétrie par rapport à la moyenne n'est pas adapté pour une largeur d'intervalle théorique, aussi bien que pour une largeur de 0,005[98] les années 2001 à 2003 où la performance moyenne des entreprises est proche de zéro (et donc où l'axe de symétrie est proche du seuil). Dans le cadre de l'étude de l'année 1997, les mesures sont limitées à des largeurs comprises entre 0,001 et 0,0341.

[98] Ce problème technique est distinct et s'ajoute au problème déjà souligné de la non symétrie observée de la courbe, mais qui n'est pas l'objet de l'étude sur l'influence de la largeur sur la mesure.

2. Évolution de la largeur des intervalles et irrégularité mesurée par moyenne

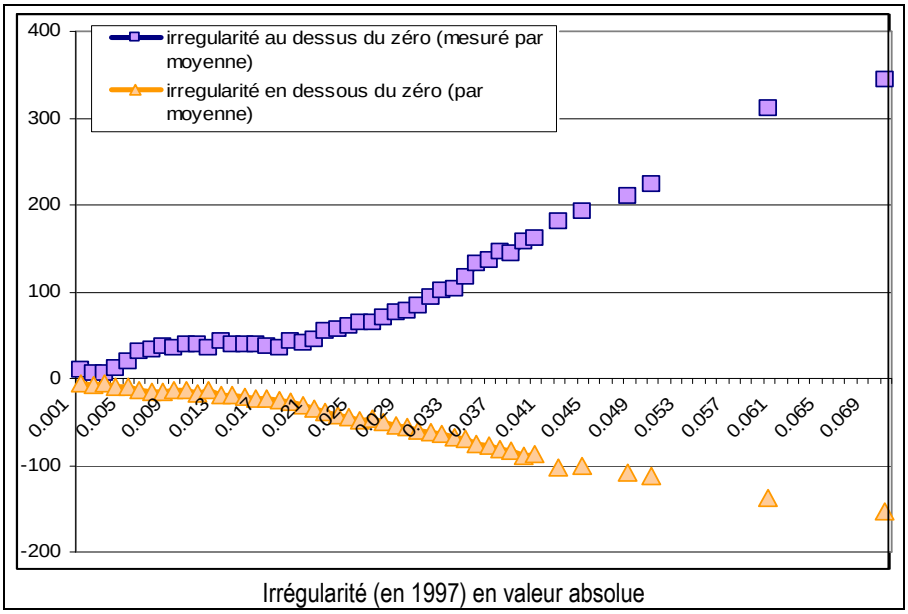
Les effectifs des deux intervalles encadrant à gauche et à droite le seuil du résultat nul sont relevés. Pour chaque largeur, 4 relevés sont ainsi effectués.

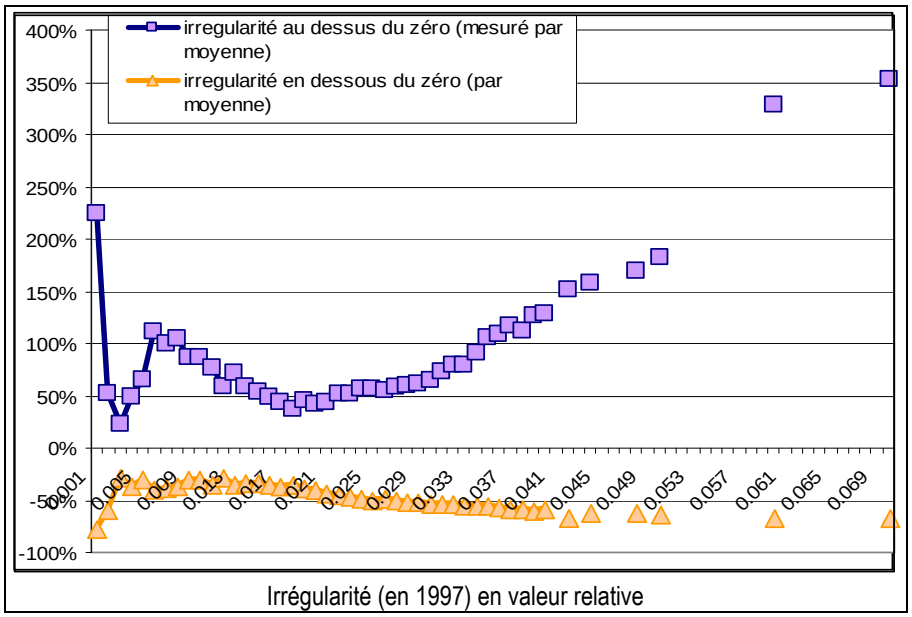
Figure 64 : Observation des 4 intervalles encadrant le seuil pour mesurer les irrégularités droite et gauche par moyenne



L'irrégularité en découle en comparant l'effectif observé et l'effectif théorique mesuré par moyenne des deux intervalles encadrant l'intervalle étudié. Les mesures sont présentées dans l'Annexe 4 : Observation des intervalles encadrant le seuil, et mesures des irrégularités par moyenne en faisant varier la largeur des intervalles. Ces résultats sont représentés graphiquement dans la Figure 65 puis commentés.

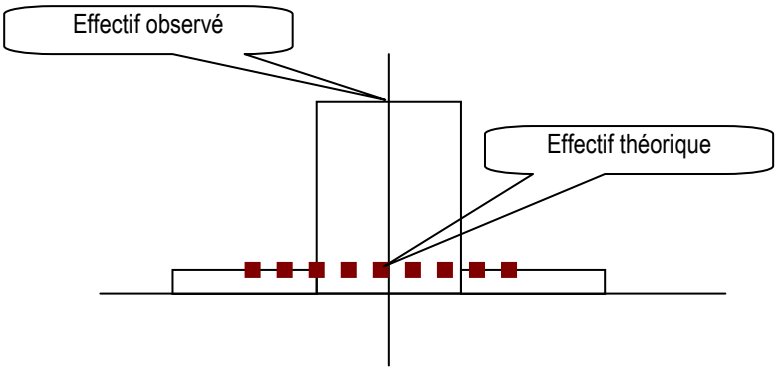
Figure 65 : Irrégularités (mesurées par moyenne en 1997) en fonction de la largeur d'intervalle





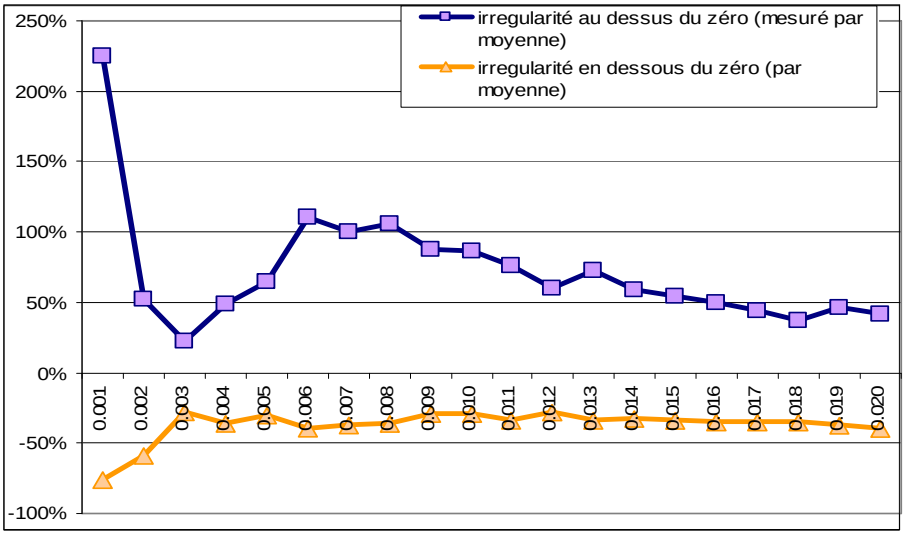
Le premier graphe présente les irrégularités en valeur absolue. Ce sont les valeurs relatives du second graphe qui sont les plus intéressantes. Il apparaît tout d'abord qu'à partir d'une certaine valeur de la largeur des intervalles, les irrégularités mesurées par moyenne deviennent extrêmement élevées. En effet, à partir d'une certaine largeur d'intervalle, l'intervalle central phagocyte à lui seul la quasi-totalité des observations. L'effectif théorique mesuré par moyenne devient alors proche de zéro, nombre qui sert de numérateur au calcul de l'irrégularité relative. La Figure 66 illustre ce propos.

Figure 66 : Lorsque la largeur est très importante, l'irrégularité tend vers l'infini car l'effectif attendu (théorique) tend vers zéro



Comme cela a été vu plus haut, les mesures effectuées pour des largeurs supérieures à 0,017 ne sont pas techniquement valides. L'étude s'est donc focalisée sur la partie du graphique présentée Figure 67.

Figure 67 : Irrégularités relatives (mesurées par moyenne) en fonction de la largeur : zoom sur les largeurs pertinentes



Il apparaît très clairement que la largeur de l'intervalle d'observation influence considérablement la mesure de l'irrégularité. L'irrégularité à droite du seuil oscille ainsi entre 23 % et 255 %. Les valeurs extrêmes sont observées sur les largeurs les plus faibles, ce qui semble logique compte tenu du faible nombre des effectifs sur de tels petits intervalles, et donc de leur plus forte volatilité.

La largeur de 0,005 la plus retenue dans la littérature (Burgstahler et Dichev, 1997 ; Burgstahler et Eames, 2003 ; Dechow, Richardson et Tuna, 2003 ; Mard, 2004 ; Brown et Caylor, 2005 ; Coppens et Peek, 2005 ; Durtschi et Easton, 2005 ; Beaver, McNichols et Nelson, 2007) conduit à mesurer une irrégularité de 65 %, mais les auteurs qui retiennent 0,01 (Leuz, Nanda et Wysocki, 2003 ; Phillips, Pincus et Rego, 2003) auraient mesuré une irrégularité de 87 %. Par ailleurs, retenir une largeur intermédiaire de 0,0025 conduit à mesurer une irrégularité proche de 25 % seulement. L'amplitude mesurée varie donc du simple au quadruple.

Par ailleurs, il est intéressant de constater que plus la largeur augmente, plus l'irrégularité diminue ce qui peut sembler cohérent avec la théorie selon laquelle les entreprises ne manipulent leurs résultats que pour dépasser de peu le seuil.

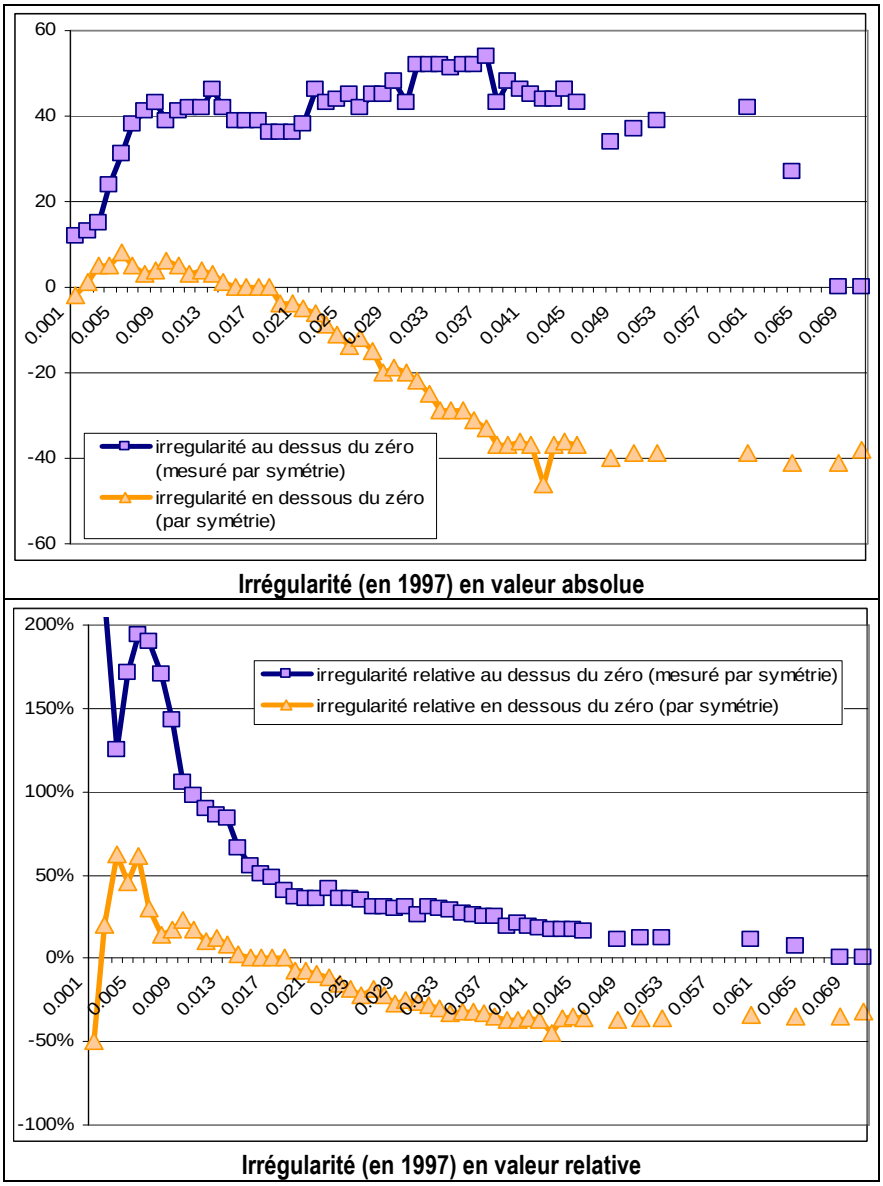
L'irrégularité à gauche du seuil, si elle est peu prononcée[99], connaît elle aussi des variations importantes (de 27 % à 77 %) lorsque la largeur de l'intervalle d'observation est modifiée.

[99] Ce constat est vrai en 1997, mais ne peut être généralisé à toutes les années.

3. Évolution de la largeur des intervalles et irrégularité mesurée par symétrie

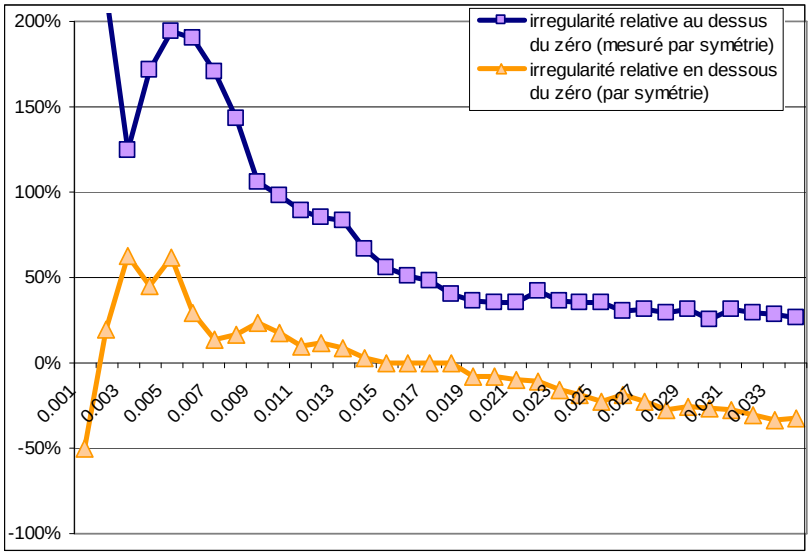
Pour estimer l'effectif théorique de l'intervalle supérieur au seuil (par exemple sur l'intervalle [0 ; 0,001]) il faut mesurer l'effectif symétriquement situé par rapport à la moyenne de la distribution (soit dans l'exemple sur l'intervalle [0,0671 ; 0,0681]). Ce procédé est ainsi mené successivement en élargissant la largeur de l'intervalle. Les mesures effectuées sont présentées dans l'Annexe 5 : Mesures des irrégularités par symétrie en faisant varier la largeur des intervalles. Les irrégularités ainsi mesurées sont illustrées graphiquement dans la Figure 68.

Figure 68 : Irrégularités (mesurées par symétrie en 1997) en fonction de la largeur d'intervalle



Les contraintes techniques exposées précédemment conduisent à focaliser l'analyse sur la partie du graphe comprise entre 0,001 et 0,034.

Figure 69 : Irrégularités relatives (en 1997) mesurée par symétrie : zoom sur les largeurs pertinentes



Sur cette figure, il apparaît clairement que la largeur de l'intervalle d'observation influence considérablement la mesure de l'irrégularité. L'irrégularité à droite oscille du simple au décuple alors que l'irrégularité à gauche change carrément de signe selon la largeur retenue.

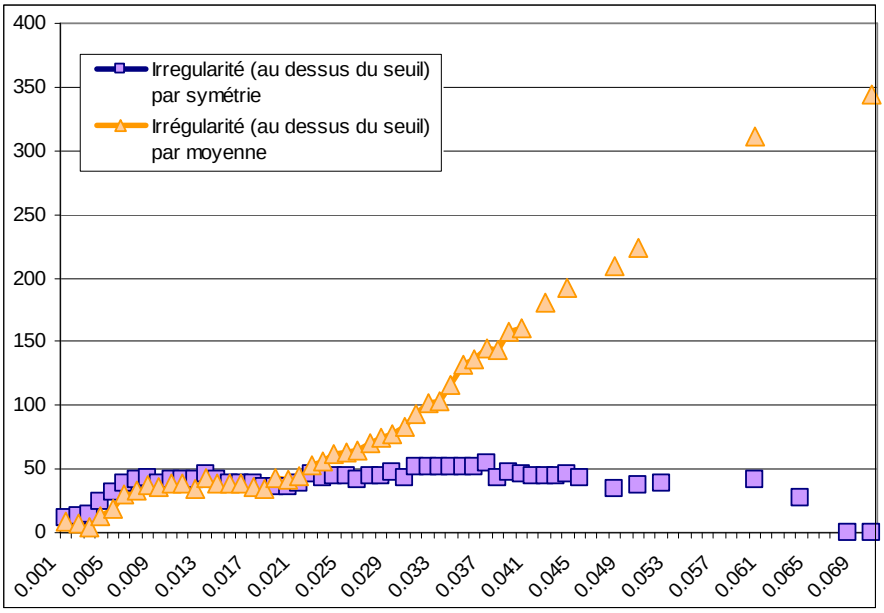
4. Comparaison des irrégularités mesurées par symétrie et moyenne

L'objet de cette partie n'est pas de comparer les deux méthodes de mesure[100], mais il est tout de même intéressant de souligner que (1) dans les deux cas, il y a une influence de la largeur sur la mesure et que (2) dans les deux cas cette influence ne se traduit pas de la même manière.

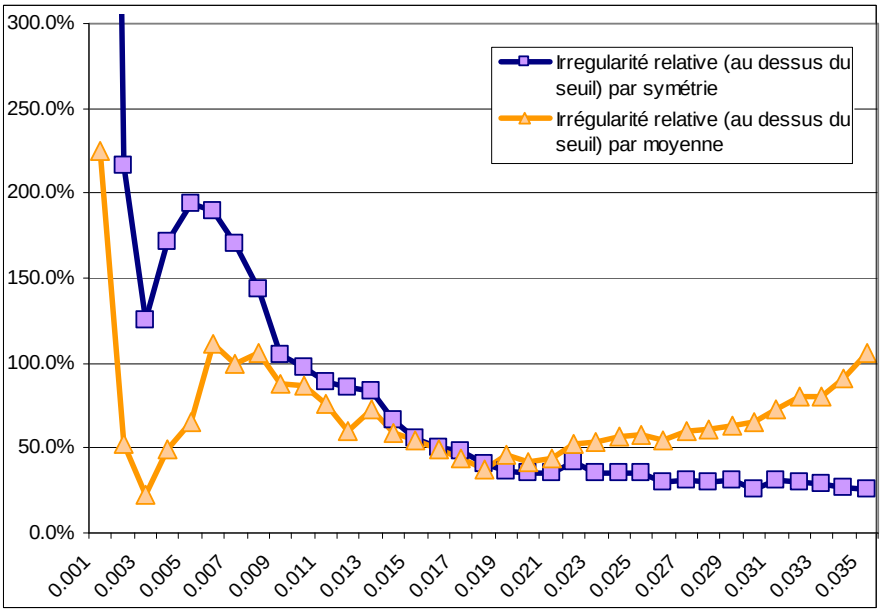
[100] Voir sur ce sujet la section A du chapitre III.

Figure 70 : Comparaison de l'influence de la largeur d'intervalle sur les mesures par symétrie et par moyenne ; irrégularité droite

Comparaison des irrégularités (relatives) mesurées par symétrie et par moyenne de l'irrégularité droite



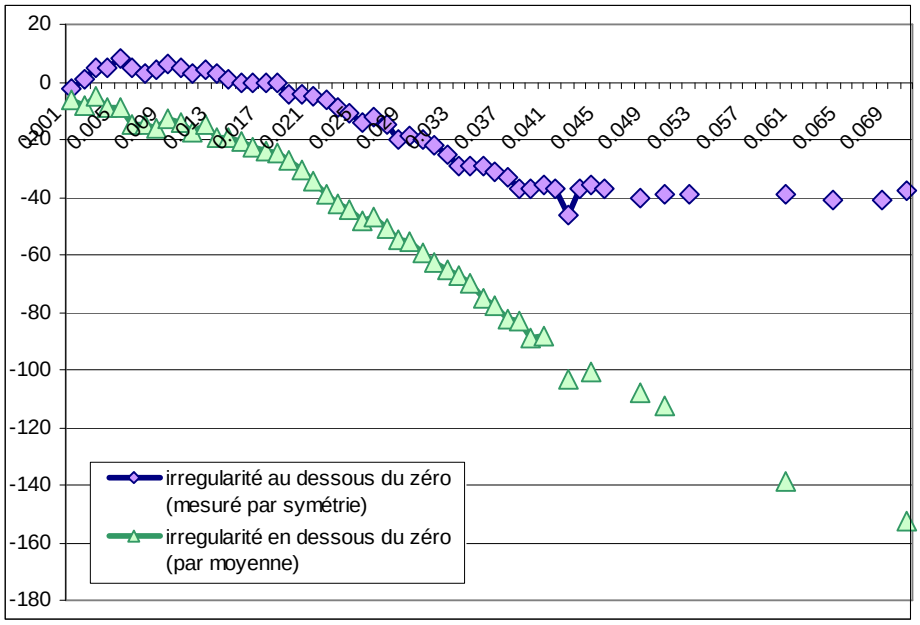
Comparaison des irrégularités (relatives) mesurées par symétrie et par moyenne de l'irrégularité droite **; zoom sur les largeurs pertinentes**



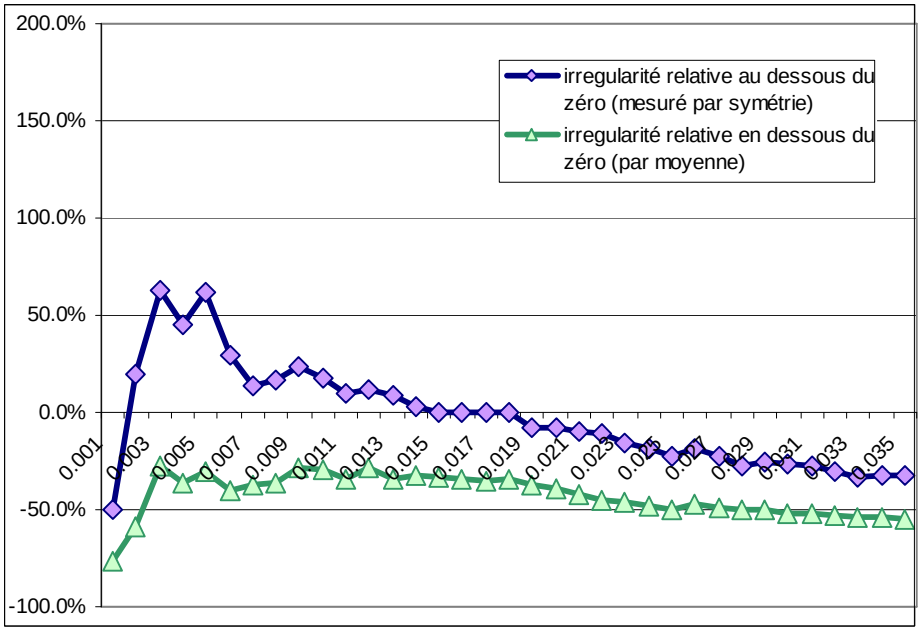
Sur les Figure 70 et Figure 71, il apparaît clairement que les tendances mesurées par symétrie et moyenne ne s'opposent pas.

Figure 71 : Comparaison de l'influence de la largeur d'intervalle sur les mesures par symétrie et par moyenne ; irrégularité gauche

Comparaison des irrégularités (relatives) mesurées par symétrie et par moyenne de l'irrégularité gauche



Comparaison des irrégularités (relatives) mesurées par symétrie et par moyenne de l'irrégularité gauche ; **zoom sur les largeurs pertinentes**



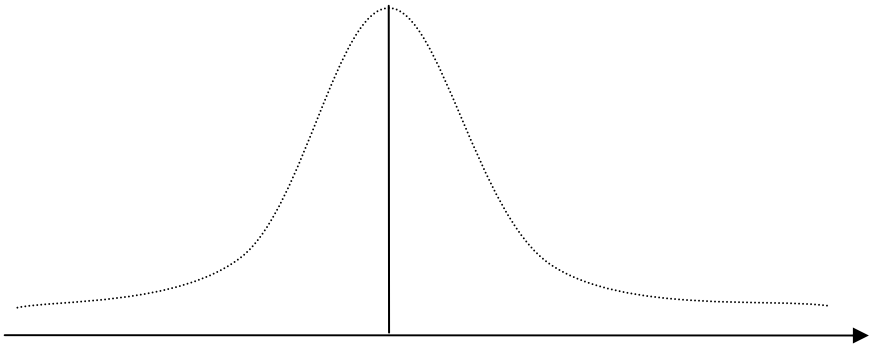
Mais il apparaît cependant que les mesures sont systématiquement et très significativement différentes. La mesure par moyenne tend vers une limite lorsque la largeur des intervalles augmente alors que la mesure par symétrie est plus volatile.

5. Proposition d'un nouvel indicateur de l'irrégularité

Les mesures de discontinuité dans la littérature académique se sont jusqu'alors limitées à la mesure de la « hauteur » de l'irrégularité, c'est-à-dire, pour une largeur d'intervalle fixée, au nombre d'entreprises sur-numéraires situées sur un intervalle observé. Or il apparaît que l'irrégularité maximale n'intervient pas, chaque année, pour la même largeur d'intervalle. Est-il alors pertinent de vouloir fixer une largeur d'intervalle unique lorsque l'objectif est de comparer des irrégularités dans l'espace et dans le temps ? Et si la mesure devait dépendre de la largeur, comment en tenir compte ?

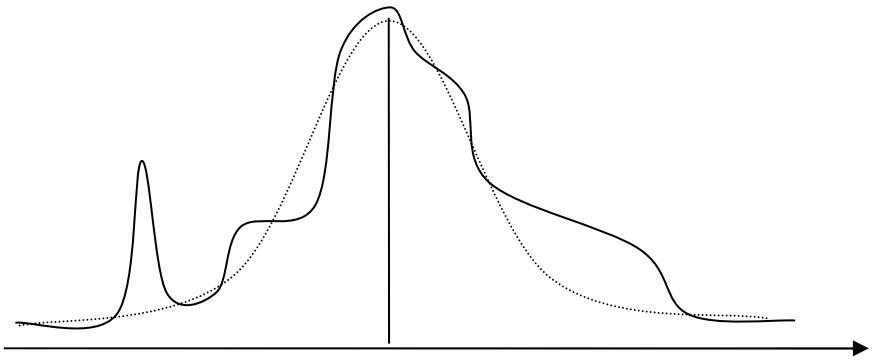
Pour répondre à cette question, il est utile de faire un petit détour explicatif. La Figure 72 représente très schématiquement une distribution théorique.

Figure 72 : Exemple de distribution



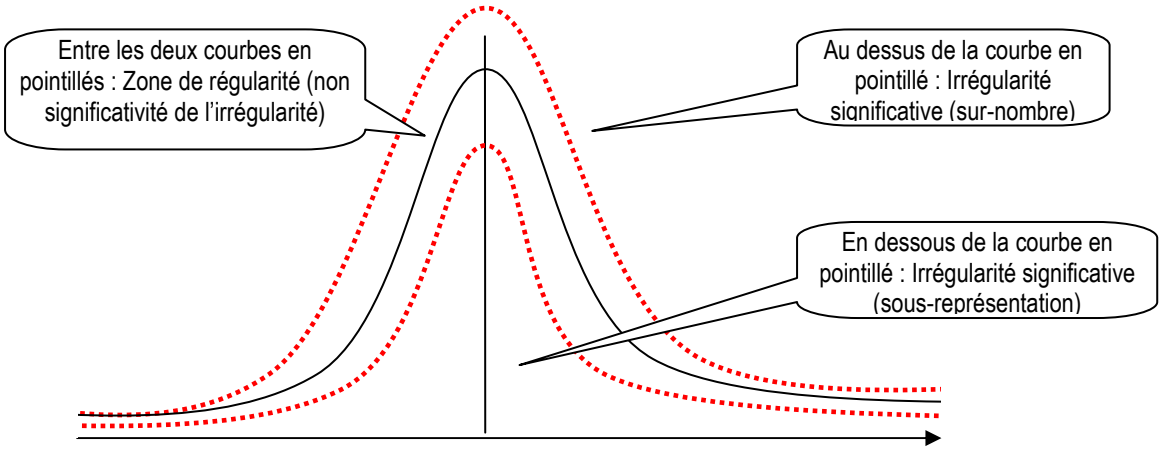
La Figure 73 reprend cette même distribution sur laquelle a été tracée une seconde distribution comportant des irrégularités.

Figure 73 : Exemples d'irrégularités



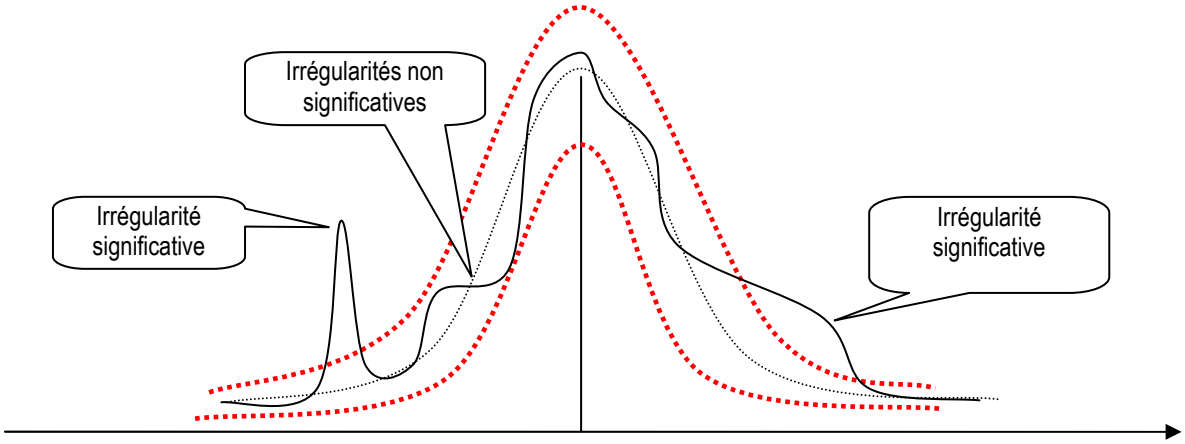
Comment savoir (graphiquement) si une irrégularité doit être considérée comme significative ou non ? Pour répondre à cette question, il est possible de dessiner une zone encadrant la distribution théorique (pointillés sur la Figure 74).

Figure 74 : Représentation graphique de la zone d'irrégularité potentielle associée à une distribution



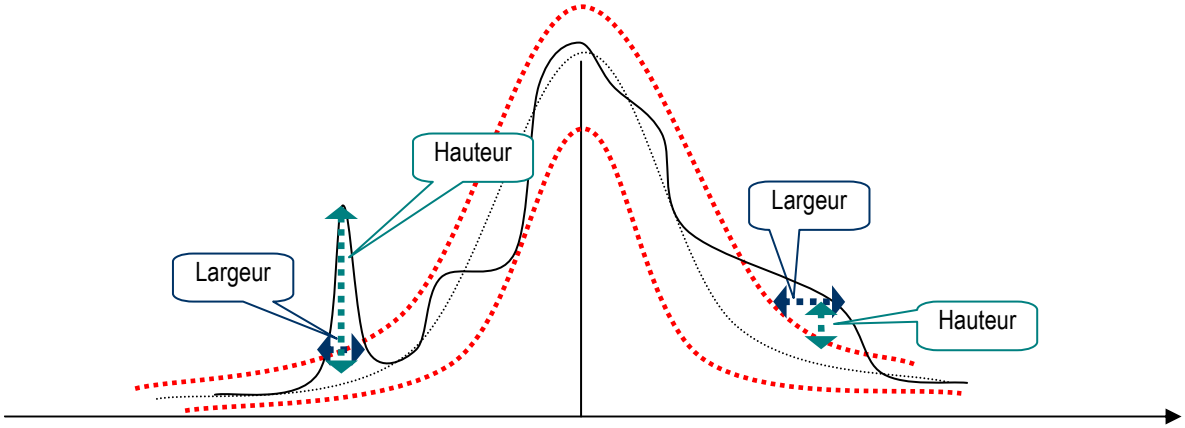
Graphiquement, il faut alors considérer que toute irrégularité qui « sort » de la zone comprise entre les deux courbes en pointillés est significative. Si l'on superpose alors les Figure 73 et Figure 74, on obtient la Figure 75.

Figure 75 : Représentation graphique de la significativité des irrégularités



Sur ce graphique, il apparaît que deux irrégularités seulement sortent de la zone de significativité. L'une est très haute et peu large, et la seconde est peu haute et très large. Graphiquement, il apparaît alors assez naturel de proposer deux critères de mesure de ces irrégularités. Une mesure de l'amplitude (« hauteur » de l'irrégularité), et une mesure de la « persistance » (« largeur » de l'irrégularité). Ces deux mesures sont illustrées graphiquement sur la Figure 76.

Figure 76 : Largeur et hauteur de l'irrégularité, deux mesures complémentaires



Dans l'exemple ci-dessus, il est évident qu'une irrégularité est plus haute et moins large, alors que la seconde est moins haute et plus large. Ne mesurer que les hauteurs des irrégularités pour une largeur fixée conduit donc à fournir une information incomplète sur leur nature. Autrement dit, il peut être utile de calculer deux indicateurs et de les comparer. Ils se complètent, un peu comme une moyenne et un écart type sont également utiles pour décrire une population. Le premier, qui mesure la « hauteur » (ou l'amplitude) de l'irrégularité à un point donné, correspond au nombre d'entreprises en trop (ou manquantes). Le second, qui mesure la « largeur » (ou la persistance) de l'irrégularité dans la distribution, donne une information sur les montants manipulés.

Appliqué à notre exemple de la Figure 76, la première irrégularité (haute et peu large) peut être interprétée comme la manifestation d'une gestion du résultat très ponctuelle, qui touche un grand nombre d'entreprises, mais qui porte sur des montants relativement faibles en valeur. La seconde correspond à une gestion du résultat qui concerne un nombre plus restreint d'entreprises, mais concernant des montants de plus grande ampleur.

Si ces deux indicateurs sont mis en évidence, des hypothèses sur les raisons de l'une ou l'autre situation peuvent être proposées. Dans le premier cas, l'impact psychologique du seuil semble très fort, mais apparemment, les entreprises sont très contraintes car elles ne peuvent pas « transférer » un fort montant de résultat. Dans la seconde situation, l'impact psychologique semble moins important, mais les entreprises ont manifestement plus de liberté pour gérer leur résultat. Les explications peuvent être réglementaires (moins de contraintes, qualité des normes, procédures d'audit, contrôle interne), liées aux motivations des dirigeants, à la structure des financements, au secteur d'activité…

Être capable de mesurer les montants manipulés, c'est répondre à l'objection faite par Mard (2004) citant Chalayer-Rouchon, Degeorge et Le Nadant (2001). Pour ces auteurs,

la méthodologie par les seuils « ne permet pas d'apprécier l'ampleur de la gestion des résultats, mais seulement de tester la présence ou non d'un comportement de gestion des résultats. » Si les mesures non-paramétriques utilisées jusqu'alors dans la littérature étaient trop sensibles aux variations de la largeur de l'intervalle pour permettre d'estimer fiablement les montants manipulés, une mesure s'affranchissant de cette limite permet de les calculer. C'est ce qui est mis en oeuvre dans la partie 3 de la thèse.

6. Conclusions et apports de l'étude de l'influence de la largeur des intervalles sur la mesure des irrégularités

1. La largeur de l'intervalle a un très fort impact sur la mesure de l'irrégularité. Cela peut conduire à inverser les résultats de certaines études (Glaum, Lichtblau et Lindemann, 2004). Cette limite est donc très préoccupante. Elle est essentiellement liée à la démarche non paramétrique de mesure mise en œuvre dans la plupart des études réalisées jusqu'alors. Ce constat conduit à imaginer la mise en place d'une démarche paramétrique qui est présentée dans la partie 3 de la thèse.

2. Le fort impact du choix de la largeur sur la mesure a comme conséquence que le choix d'une largeur ad hoc n'est pas adapté à des études comparatives entre populations dont les caractéristiques (taille, sommet de la distribution …) sont différentes. L'utilisation d'une largeur théorique semble donc souhaitable. Dans le cadre de l'étude de l'année 1997, utiliser la largeur théorique de Silverman (1986) et Scott (1992) semble fournir une mesure médiane de l'irrégularité, ce qui n'est pas le cas des largeurs ad hoc. Ce résultat mériterait d'être confirmé sur plusieurs années avant d'être généralisé.

3. Les mesures par symétrie comme par moyenne rencontrent des limites techniques qui interdisent de mesurer l'irrégularité pour certaines largeurs d'intervalle (au delà de m[101] pour les mesures par symétrie, et de m/2 pour les mesures par moyenne), et interdisent l'étude des années où sommet de la distribution et seuil sont proches l'un de l'autre.

[101] Avec m = axe de symétrie de la distribution (médiane ou moyenne).

4. Enfin, l'influence de la largeur des intervalles sur les mesures conduit à s'interroger sur l'intérêt de définir une largeur fixe des intervalles. Au contraire, faire varier la largeur permet de maximiser l'irrégularité, démarche plus pertinente dans l'objectif d'identifier la totalité des entreprises susceptibles de manipuler leurs comptes. La largeur de l'intervalle n'est alors plus considérée comme une contrainte, mais comme un indicateur mesurant une nouvelle dimension de l'irrégularité : le montant manipulé.

En conclusion, ces constats conduisent à relativiser toutes les mesures d'irrégularité non paramétriques qui nécessitent, par construction, de fixer une largeur d'intervalle. Seules des mesures paramétriques permettent de s'affranchir de l'impact de la largeur de l'intervalle.

C. Impact du choix de la variable de mise à l'échelle

Lorsque l'on étudie la distribution d'une variable réduite (c'est-à-dire divisée par un indicateur de taille), les irrégularités observées peuvent provenir d'une irrégularité affectant le numérateur, ou d'une irrégularité affectant le dénominateur. Jusque là, l'impact du dénominateur a été ignoré, comme le font la grande majorité des travaux en comptabilité sur le sujet. Mais ce problème méthodologique, soulevé par Durtschi et Easton (2005), Dechow, Richardson et Tuna (2003) et McNichols (2003), est crucial puisque si l'irrégularité constatée est due au dénominateur, elle remet alors en cause les conclusions faites sur la présomption de manipulation du résultat. Le travail sur l'impact du dénominateur sur les seuils comptables se décompose en deux questions :

1. Existe-t-il une irrégularité sur la distribution de la variable résultat non réduit ? S'il existe une irrégularité sur la variable non réduite, alors l'impact du dénominateur joue éventuellement sur sa mesure, mais ne remet pas en cause son existence.

2. Est-il neutre d'utiliser un indicateur de taille plutôt qu'un autre ? Dans la littérature, trois indicateurs de taille ont été utilisés pour mettre le résultat à l'échelle : le total actif (AT), le chiffre d'affaires (CA) et la valeur de marché (VM)[102].

Ces deux questions structurent la section C.

1. Distribution des résultats non réduits

Les paramètres de la distribution du résultat net (IB[103]) non réduit (non « mis à l'échelle » par un indicateur de taille) dans la base sont présentés dans le Tableau 40.

[102] Dans la base Global, AT correspond à « Assets Total », CA est « Sales », et VM est MV (« Market Value »). Dans la thèse, les abréviations françaises sont privilégiées, mais dans certains tableaux, les abréviations anglaises peuvent demeurer.

[103] IB (Income Before extraordinary items) est le résultat net dans la base Global Vantage.

Tableau 40 : paramètres des distributions du résultat net non mis à l'échelle

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Agrégation sur 13 ans
Nombre	291	296	376	423	617	702	754	798	791	734	698	659	611	7750
Moyenne	41.840	21.080	28.133	16.668	27.359	44.271	58.337	65.142	97.570	19.450	-4.900	65.171	105.393	48.432
Min	-664.048	-1252	-1815	-3886	-2641	-1116	-918	-970	-349	-13597	-23301	-2165	-3610	-23301
Premier quartile	0.785	0.240	1.372	0.670	0.906	0.813	0.850	0.661	0.448	-0.449	-1.156	-0.575	0.105	
Médiane	9.701	7.823	8.879	7.017	4.747	4.587	4.288	3.795	3.796	3.092	2.517	2.651	4.234	4.405
Troisième quartile	40.232	42.306	42.168	37.021	26.455	23.320	20.671	21.847	23.579	20.104	18.451	21.683	28.977	
Max	928	1032	1486	1396	1141	2402	2442	2903	6904	7658	5941	7025	9612	9612
écart type	145	187	206	296	193	193	228	260	488	736	1247	463	592	538.467
IQR	39.448	42.066	40.796	36.350	25.549	22.508	19.821	21.186	23.131	20.553	19.607	22.258	28.872	24.828
2(IQR)n-1/3	11.906	12.624	11.305	9.685	6.002	5.065	4.355	4.568	5.002	4.557	4.421	5.116	6.805	2.509

Les distributions ont été tracées en retenant les paramètres suivants :

- largeur d'intervalle d'observation unique et arbitraire égale à 1,25. Une largeur arbitraire égale pour toutes les distributions est retenue pour faciliter les comparaisons graphiques. C'est un chiffre arrondi qui permet de fixer facilement un point de départ[104] de la distribution permettant de faire correspondre le seuil du résultat nul à une frontière entre deux intervalles. Cette valeur est inférieure (la moitié à un arrondi près) à la largeur théorique d'observation sur la base agrégée (2,509). Elle correspond également (toujours à un arrondi près) à la moitié de la plus petite médiane (2,517 en 2002). Les moyennes étant influencées par les valeurs extrêmes, c'est en effet la médiane qui est retenue comme sommet des distributions. Il est donc utile de sélectionner une largeur inférieure à la médiane afin de distinguer clairement les effectifs qui encadrent la médiane de ceux qui encadrent le zéro. Une largeur de 1,25 assure la présence de deux intervalles au moins entre les deux.

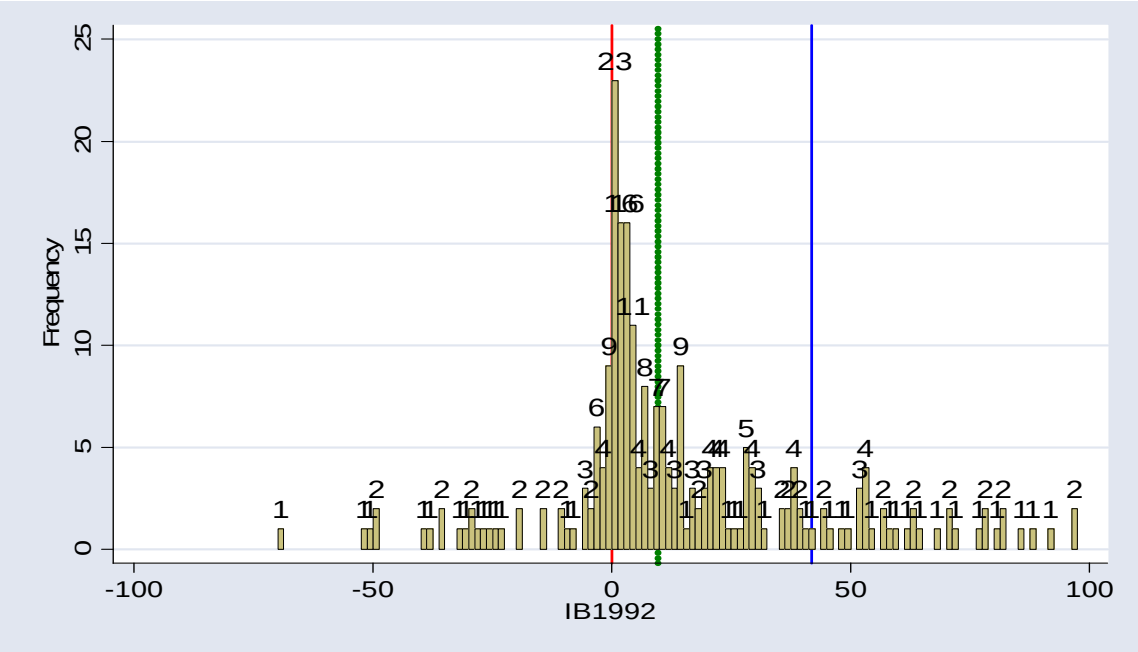
- Les graphes font un « zoom » sur la partie centrale de la distribution, soit entre les valeurs]-100;+100[. Cet intervalle représente 6704 observations sur les 13 années, soit 86,5 % des observations. Le point de départ est fixé à −100,0000000001 (neuf chiffres après la virgule au moins) pour s'assurer que les valeurs strictement égales à zéro figurent sur l'intervalle [0 ;+1,25[. Seule l'allure sur la partie centrale est intéressante dans le cadre de cette étude.

[104] Commande « start » sous Stata.

- La médiane est tracée en pointillé gras (vert), la moyenne en continu (bleu). Le résultat nul est tracé en pointillé (rouge).

À titre d'illustration, la Figure 77 ne représente que la distribution des résultats de l'année 1992. Les 13 distributions sont présentées en Annexe 6 : Distributions annuelles de la variable résultat non mise à l'échelle.

Figure 77 : Distribution de la variable résultat non mise à l'échelle en 1992



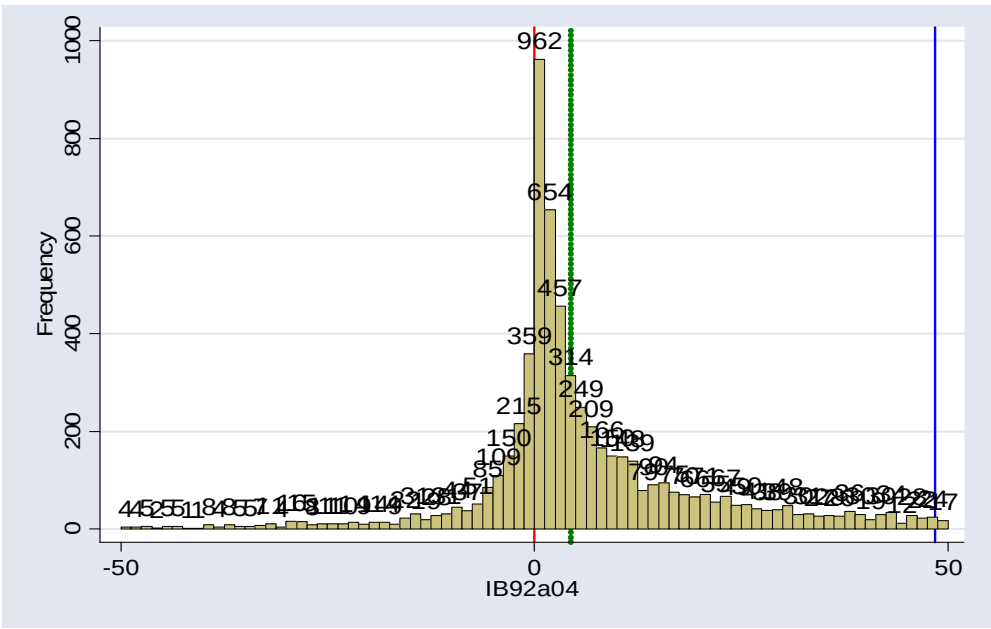
L'observation des 13 distributions des résultats non réduits par un indicateur de taille fait apparaître clairement deux phénomènes : (1) les sommets des distributions sont tous situés sur l'intervalle juste à droite du zéro et (2) l'effectif des entreprises faiblement déficitaires est toujours très nettement inférieur à l'effectif des entreprises faiblement bénéficiaires.

La distribution des résultats agrégés (comme le font les auteurs en comptabilité[105]) sur 13 années est ensuite tracée[106].

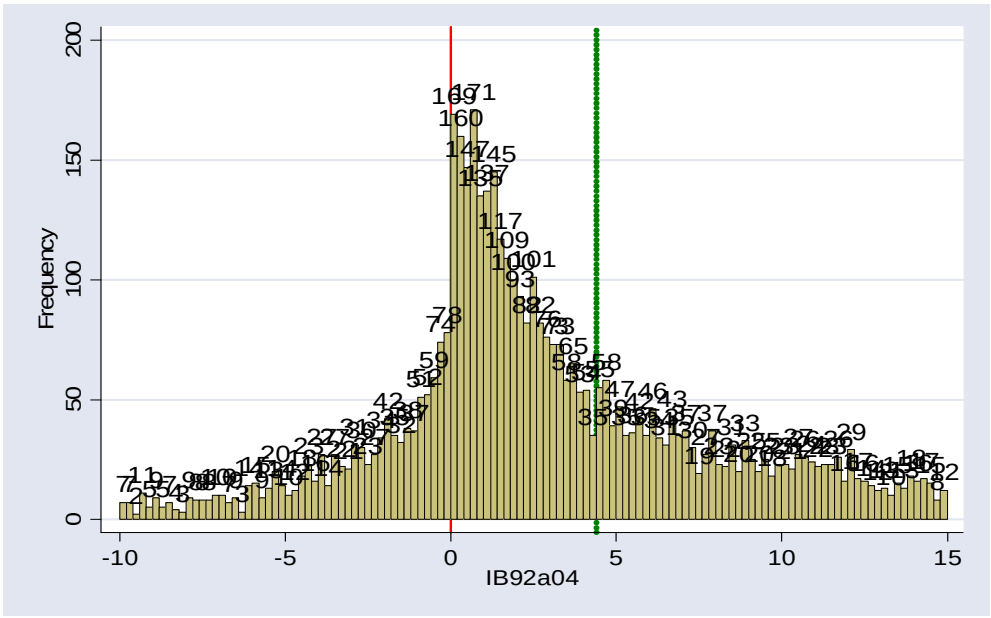
[105] Cette pratique est discutée dans la partie 3 de la thèse.

[106] Pour ce faire, la variable combinée (fonction « stack » sous stata) IB92a04 est créée.

Figure 78 : Distribution de la variable résultat (non mise à l'échelle) agrégée sur 13 ans



histogram IB92a04 if IB92a04>-50 & IB92a04<50, width(1.25) start(-50.0000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(48.432, lcolor(blue)) xline(4.405,lcolor(green) lwidth(thick) lpattern(dot))



histogram IB92a04 if IB92a04>-10 & IB92a04<10, width(0.2) start(-10.0000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(4.405,lcolor(green) lwidth(thick) lpattern(dot))

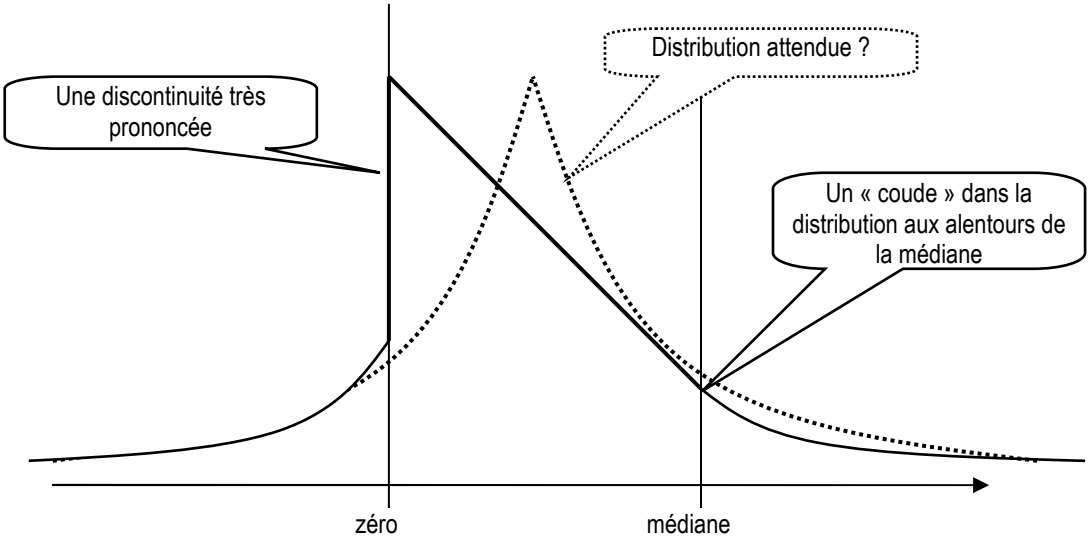
Comme l'observation des 13 distributions annuelles l'avait déjà montré, cette distribution se caractérise par (1) une concentration autour de l'intervalle supérieur au résultat nul (où se situe le sommet de la distribution, c'est-à-dire le mode)[107] et (2) une forte discontinuité autour du seuil du résultat nul.

Par ailleurs, l'observation de cette distribution (plus lisse car le nombre d'observations est plus important) conduit à observer un « coude » dans l'allure de la distribution autour de la

[107] L'allure de la distribution n'est en aucun cas proche d'une allure « Normale ». Elle est beaucoup trop pointue (kurtosis = 841) et dissymétrique (skewness = -17,93). Ce sujet est discuté en partie 3.

médiane. Autrement dit, les observations situées entre la médiane et le seuil du résultat nul forment un « triangle » inattendu. Ce constat est schématisé dans la Figure 79.

Figure 79 : Allure apparente de la distribution des résultats non mis à l'échelle



Dès lors, il est possible de conclure qu'il y a manifestement une irrégularité dans la distribution des résultats bruts autour du seuil du résultat nul. En effet, le résultat nul est un point invariant qui ne dépend pas des paramètres de la distribution annuelle (comme la moyenne ou la médiane). Il n'y a donc aucune raison économique que le sommet se situe systématiquement juste au dessus de ce seuil. Par ailleurs, l'importante asymétrie entre l'effectif juste inférieur au seuil et celui juste supérieur renforce ce constat (la courbe n'est pas centrée « sur » le résultat nul).

Conclusion et apports sur la distribution des résultats non réduits

En conclusion, l'observation de la distribution des résultats non mis à l'échelle par un critère de taille conduit à confirmer l'existence d'une irrégularité au seuil du résultat nul. Cette irrégularité ne peut en aucun cas être imputée uniquement à une irrégularité de distribution du dénominateur. Ce constat est important car il renforce la présomption de gestion du résultat pour éviter un seuil.

Ceci étant dit, il n'en demeure pas moins que l'étude des pratiques de gestion du résultat pour éviter le seuil du résultat nul ne peut s'effectuer sur la seule variable résultat. Tout d'abord parce que demeure le problème de mise à l'échelle : À partir de quand peut-on dire qu'un résultat est « faible » ou « proche de zéro » ? Un petit résultat pour une grande entreprise peut être considéré comme un grand résultat pour une petite entreprise. La mise à l'échelle vise à relativiser l'effet taille. Par ailleurs, l'irrégularité qui apparaît autour

du résultat nul étant toujours située au sommet de la courbe, cela rend délicat toute tentative de mesure. Seule la méthode par extrapolation est utilisable, mais il a déjà été souligné dans la section B du chapitre V que ce calcul dépend fortement de l'allure attendue en ce point (exponentielle, logarithmique, linéaire). La mise à l'échelle permet de répondre à ces deux problèmes.

2. Comparaison des variables de mise à l'échelle

Trois différents dénominateurs ont été utilisés dans la littérature pour mettre à l'échelle les résultats publiés (AT, CA et VM). Dans une première sous partie (a), les données de la base concernant chacun de ces indicateurs sont examinées. Ensuite, l'impact de ces trois dénominateurs est comparé en deux temps. Dans un premier temps (b), seules les mesures effectuées sur les observations communes sont comparées, c'est-à-dire les entreprises/année pour lesquelles sont disponibles le résultat net, l'actif total, le chiffre d'affaires, et la valeur de marché. Dans un deuxième temps (c) la comparaison est étendue à l'ensemble de la population afin de mettre en évidence l'impact du biais de sélection généré par le choix d'un dénominateur.

a. Les variables de mise à l'échelle

Les dénominateurs utilisés dans la littérature sont : AT (Actif Total), CA (Chiffre d'Affaires) ou VM (Valeur de Marché).

a.1. L'actif total

L'actif total (AT) est le dénominateur retenu dans le cadre des études menées dans les sections A et B sur l'impact du choix de la méthode de mesure, et sur l'impact de la largeur d'intervalle. Ce dénominateur a été présenté dans la section C du chapitre IV. La base ayant été complétée (voir chapitre V), le taux d'observations communes entre les variables IB et AT est proche de 100 % (voir Tableau 41).

Tableau 41 : Comparaison du nombre d'observations de la variable AT et de la variable IB dans la base

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
IB	291	296	376	423	617	702	754	798	791	734	698	659	611	7750
AT	292	296	376	422	617	704	755	802	791	737	699	660	611	7762
AT mais pas IB	1	0	0	0	1	2	2	4	0	3	1	1	1	16
IB mais pas AT	0	0	0	1	1	0	1	0	0	0	0	0	1	4
IB et AT	291	296	376	422	616	702	753	798	791	734	698	659	610	7746
														Soit 99.95%

Comme l'actif total inclut par construction le résultat annuel (IB), il doit être transformé pour que le dénominateur soit indépendant du numérateur. Deux transformations sont proposées ; $AT_{(n-1)}$ et (AT-IB). La section C du chapitre IV présente les avantages théoriques de la transformation (AT-IB). L'étude de la base de données menée dans le chapitre V montre que cette transformation permet d'éviter la perte de près de 13 % des données, et évite les valeurs extrêmes. Dans le cadre de cette sous-partie, une étude complémentaire est menée afin de contrôler empiriquement dans quelle mesure l'utilisation de IB/(AT-IB), dont la distribution est définie sur [-1 ; +∞[, ne modifie pas l'étude des seuils par rapport à l'utilisation de la variable $AT_{(n-1)}$ dont la distribution est, elle, définie sur [-∞ ; +∞[. Dans un premier temps, les caractéristiques de chacune des deux distributions sont comparées (en accordant une importance particulière à la partie centrale, où se situe le seuil). Puis les observations situées à proximité du seuil (sur les deux intervalles encadrant le seuil) sont comparées.

a.1.1. Caractéristiques des distributions des variables IB/$AT_{(n-1)}$ et IB/(AT-IB)

Pour comparer les distributions selon les deux transformations du dénominateur actif total, les médianes, quartiles, kurtosis et skewness sont comparés. Les distributions sont tracées, ainsi que les diagrammes en boites à moustaches.

Tableau 42 : Comparaison des distributions des deux transformations de la variable AT

stats	Ib92a04	RsurAmR[108]	RsurAm1	Commentaires
N	7750	7746	6735	L'utilisation de $AT_{(n-1)}$ => perte de près de 13%
min	-23301	-0.9809082	-4.322242	
max	9612	6.817411	7.214285	
mean	48.43211	0.0229601	0.0174588	Les indicateurs de tendance sont très proches
P50	4.405278	0.0263726	0.0273305	
sd	538.4668	0.1363194	0.1942774	
kurtosis	841.5353	823.3709	557.5533	Réduire par $AT_{(n-1)}$ aplatit la courbe
skewness	-17.93202	16.85091	9.789404	Les distributions sont très asymétriques vers la droite. L'asymétrie semble moins prononcée en utilisant $AT_{(n-1)}$

La lecture du tableau ci-dessus nous montre que l'utilisation de $AT_{(n-1)}$ conduit à perdre près de 13% de la base. Les indicateurs de tendance sont très proches dans les deux transformations. Les distributions sont toutes très pointues, mais réduire par $AT_{(n-1)}$ semble aplatir la courbe (kurtosis inférieur). Les deux transformations conduisent à des distributions très asymétriques vers la droite même si l'asymétrie semble moins prononcée en utilisant $AT_{(n-1)}$. Mais quelques valeurs extrêmes influencent énormément cet indicateur. C'est pourquoi les parties centrales des distributions (sur l'intervalle [-1 ; 1[) sont ensuite étudiées dans le Tableau 43.

Tableau 43 : Comparaison des deux transformations de la variable AT sur l'intervalle [-1 ;1[

stats	ib92a04	RsurAmR	RsurAm1	Commentaires
N	6701	6701	6701	Seules les observations pour lesquelles IB / $AT_{(-1)}$ sont comprises sur l'intervalle [-1 ; 1[sont retenues
min	-23301	-0.7409929	-0.9528921	
max	7658	0.8013707	0.7722842	
mean	43.27463	0.0236174	0.0204622	
p50	4.479922	0.0265588	0.0274037	
Sd	541.9096	0.0941587	0.112533	
kurtosis	927.4541	14.90227	17.58974	Les deux transformations présentent une concentration élevée
skewness	-21.69907	-1.351214	-1.939454	L'ordre de grandeur de l'asymétrie est similaire

En « zoomant » sur la partie centrale de la distribution, les deux distributions présentent toujours une concentration comparable, beaucoup plus élevée que la distribution Normale. Par contre, l'asymétrie change de signe. Cette différence est donc essentiellement liée à l'impact des valeurs extrêmes. Un nouveau « zoom » est donc effectué sur l'intervalle (sur l'intervalle [−0,5 ; 0,5[).

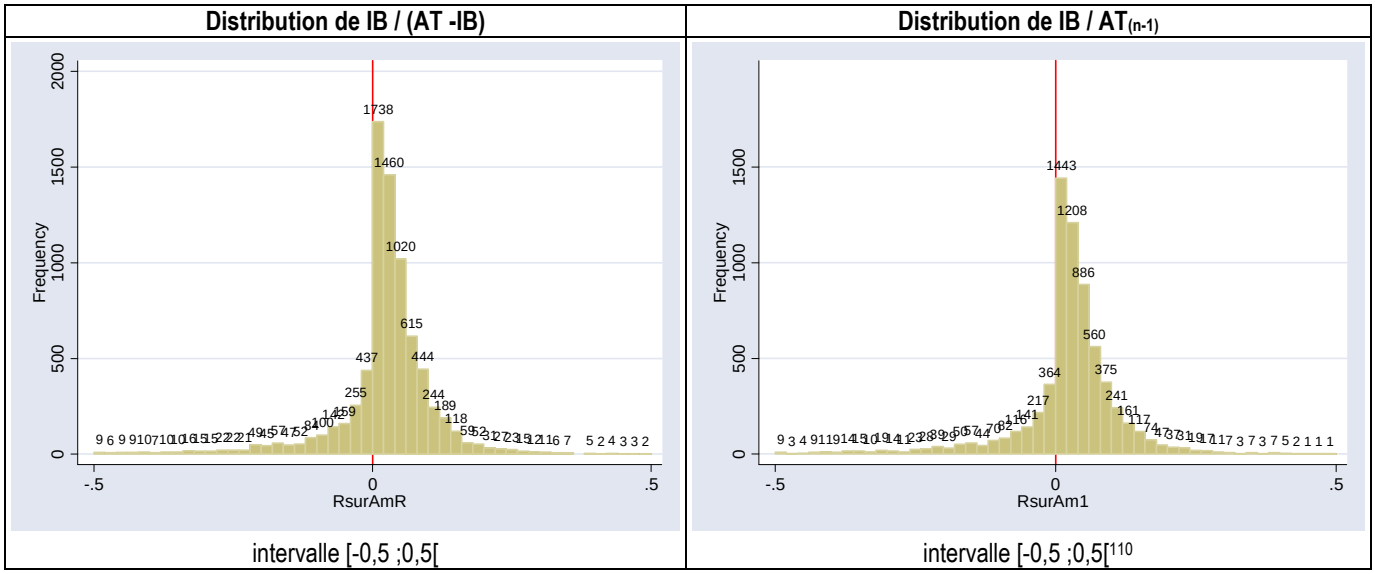
[108] Sous Stata, la variable dénommée RsurAmR correspond à IB/(AT-IB) et RsurAm1 correspond à IB/$AT_{(n-1)}$.

Tableau 44 : Comparaison des deux transformations de la variable AT sur l'intervalle [−0,5 ;0,5[[109]

Stats	RsurAm1	RsurAmR
N	6626	6626
Max	0.4811238	0.484375
Min	-0.4991482	-0.4977242
Mean	0.02385	0.0257611
P50	0.0277744	0.0268749
Sd	0.0928891	0.0834687
Skewness	-1.273286	-0.9958667
Kurtosis	9.752043	10.4589

L'étude de la partie centrale de la distribution montre que, quel que soit le dénominateur, l'allure est très similaire. Skewness et Kurtosis sont très proches. C'est ce que l'on observe graphiquement sur la Figure 80 qui trace les parties centrales de la distribution des deux transformations de la variable AT.

Figure 80 : Comparaison graphique des distributions selon les deux transformations de la variable AT ; « zoom » sur la partie centrale des distributions



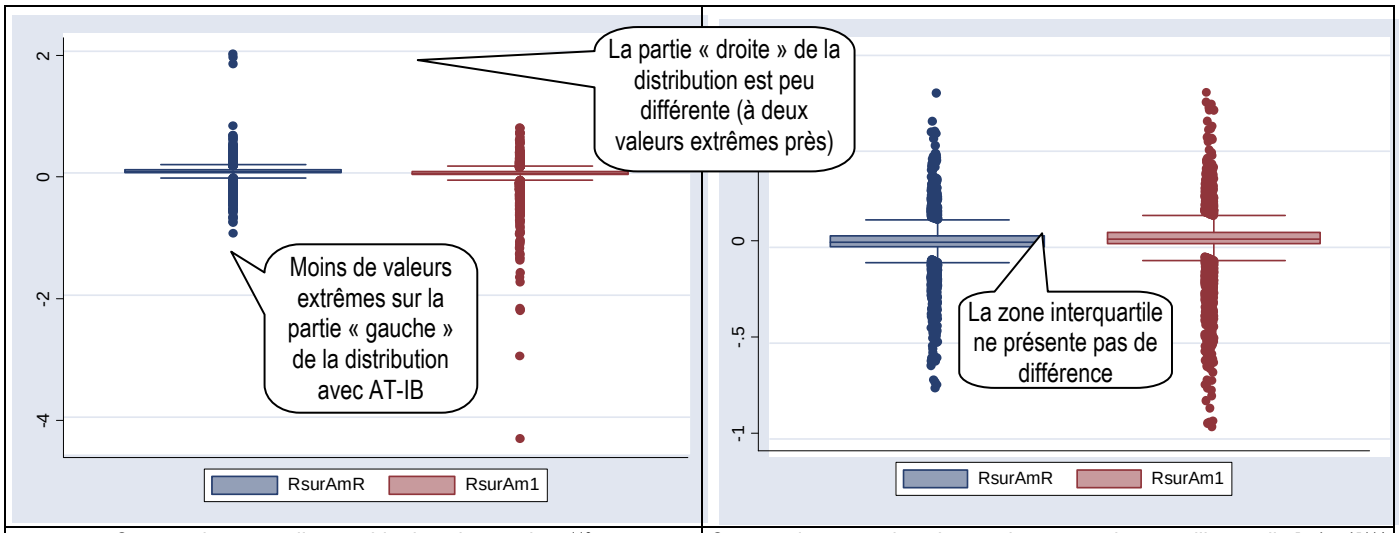
Ce qu'il importe de retenir dans ces deux graphes, c'est qu'aucune différence flagrante dans l'allure n'est constatée visuellement, que l'on divise par $AT_{(-1)}$ ou par AT-IB. La dissymétrie et la forte concentration sont toujours présentes. Il apparaît la même forme en « entonnoir inversé ».

[109] Fonction sous Stata : tabstat RsurAm1 RsurAmR if RsurAm1>-0.5 & RsurAm1<0.5 & RsurAmR>-0.5 & RsurAmR<0.5 , statistics(count max min mean median sd skewness kurtosis) columns (variables)

[110] Fonction sous Stata : "histogram RsurAm1 if RsurAm1>-0.5 & RsurAm1<0.5, start(-0.5000000001) width(0.02) frequency addlabels addlabopts(mlabcolor(black)) xline(0,lcolor(red)) yscale(range(. 2000))"

Pour mesurer l'impact des valeurs extrêmes, les diagrammes en boites à moustaches[111] ont été tracées pour chacune des transformations (Figure 81).

Figure 81 : Diagrammes en boites à moustaches[112] des deux transformations de la variable AT



Comparaison pour l'ensemble des observations[113]	Comparaison pour les observations comprises sur l'intervalle]−1 ; 1][114]

Ces diagrammes font clairement apparaître la similitude des distributions. Les seules différences sont liées aux valeurs extrêmes. Par construction, le dénominateur A-R permet de regrouper la queue gauche de distribution, et permet donc de diminuer le nombre de valeurs extrêmes négatives sans pour autant supprimer ces valeurs. La queue droite de la distribution est dans les deux cas comparativement moins étendue. Le dénominateur (AT-IB) apparaît donc comme étant un moyen de réduire la dispersion tout en conservant les données perdues lorsque l'on utilise l'actif de l'année précédente. En conclusion, la partie centrale de la distribution étant l'objet de l'étude, l'utilisation de (AT-IB) comme dénominateur permet d'éviter certaines des propriétés mathématiques indésirables de $AT_{(n-1)}$ sans biais manifeste.

[111] Ces diagrammes très simples résument essentiellement la position de la médiane, des quartiles, et des valeurs extrêmes. Un rectangle rejoint les premiers et troisièmes quartiles, à l'intérieur duquel la médiane figure par un trait. Deux traits sont tracés à 1,5 fois l'intervalle interquartile au-delà du premier et du troisième quartiles. Au-delà de ces traits, les valeurs extrêmes sont figurées par des points.

[112] En anglais : "Box plot". Sous Stata, la fonction s'appelle "graph box".

[113] Fonction sous Stata : graph box RsurAmR RsurAm1 if RsurAmR<5 & RsurAm1<5, medtype(line)

[114] Fonction sous Stata : graph box RsurAmR RsurAm1 if RsurAmR>-1 & RsurAmR<1 & RsurAm1>-1 & RsurAm1<1, medtype(line)

a.1.2. Comparaison des observations à proximité du seuils

Pour s'assurer de la pertinence de l'utilisation du ratio IB/(AT-IB), les entreprises faiblement bénéficiaires (déficitaires) selon ce ratio sont comparées aux entreprises faiblement bénéficiaires (déficitaires) selon $IB/AT_{(n-1)}$. Pour effectuer cette comparaison, les entreprises dont le résultat réduit est compris entre -0,001 et 0,001 (perte inférieure à 0,1 % et bénéfice inférieur à 0,1 % de l'actif) sont listées dans le Tableau 45.

Tableau 45 : Comparaison du nombre d'entreprises faiblement performantes d'après les variables $IB/AT_{(n-1)}$ et IB/(AT-IB)

Nombre de cas	122	
Nombre de cas identifiés par les deux méthodes		92
Nombre de cas identifiés par une méthode et pas l'autre		30
Nombre de cas identifiés uniquement par AT(n-1)	4	
Nombre de cas identifiés uniquement par AT-IB	26	Dont 11 Cas en 1992, et 8 entrées dans la base soit 19 cas indétectables par AT(n-1)
Nombre de cas identifiés uniquement par AT(n-1)	**4**	**11 cas différents sur une base équivalente, soit[115] 9%**
Nombre de cas identifiés uniquement par AT-IB	**7**	

91 % des observations sont identiques avec les deux méthodes sur des périodes d'observation comparables. La méthode (AT-IB) permet cependant de détecter davantage de cas (près de 15,6[116] %) du seul fait qu'elle intègre une année d'observation supplémentaire. Les 11 cas différents sont comparés (Tableau 46) pour mesurer l'importance des écarts.

[115] (7+4)/122 = 9%

[116] 19/122 = 15,6%

Tableau 46 : Entreprises faiblement performantes identifiées par un seul des deux dénominateurs

	Année	IB/AT-1	IB/(AT-IB)	différence	En %	Commentaire
CIE FRANCAISE MOULAGES METAL	2003	0.00083	0.00100	0.00017	16.77%	Aucune différence ne dépasse 100% ; Cela signifie que toutes les entreprises sont identifiées comme faiblement performantes par les deux dénominateurs à un intervalle près.
MECELEC	2004	0.00099	0.00109	0.00010	8.96%	
BAIL INVESTISSEMENT FONCIERE	1998	0.00095	0.00108	0.00013	11.68%	
CIE PARISIENNE DE REESCOMPTE	1998	-0.00099	-0.00151	-0.00051	34.08%	
CASINO MUNICIPAL DE CANNES	1995	-0.00100	-0.00648	-0.00548	84.55%	
EUROTUNNEL SA	1999	0.00105	0.00077	-0.00028	-36.40%	
NATEXIS BANQUES POPULAIRES	1997	0.00100	0.00096	-0.00004	-4.37%	
CREDIT AGRICOLE INDOSUEZ	1997	0.00136	0.00075	-0.00061	-81.25%	
FINAXA	1995	0.00103	0.00084	-0.00019	-23.33%	
ENTENIAL	1998	0.00142	0.00083	-0.00059	-70.34%	
MAITRE FOURNIL SA	2000	0.00135	0.00089	-0.00046	-51.78%	

Les 11 différences identifiées sont toujours de très faible ampleur. Dans 4 cas, à une question d'arrondi près, les différences sont quasi nulles. Dans aucun cas, la différence ne dépasse 100 %. Autrement dit, les sociétés détectées sur un intervalle d'observation par une méthode sont toujours situées, dans le pire des cas, sur un intervalle contiguë par l'autre méthode. Il en découle que l'utilisation du ratio IB/(AT-IB) ne semble pas fournir de résultats différents par rapport à l'utilisation du ratio $IB/AT_{(n-1)}$.

En conclusion, l'utilisation du ratio IB/(AT-IB) semble être à privilégier pour trois raisons :

1. Il a l'avantage d'augmenter très sensiblement le nombre d'observations. La perte de la première année pourrait laisser penser que l'on ne gagne qu'un treizième des données (soit 7,7 %). En réalité, la première observation de chaque entreprise entrant dans la base est « gagnée », soit 13,05 %.

2. Il limite par construction le nombre de cas extrêmes négatifs sans pour autant supprimer ces observations, ce qui n'a pas d'influence sur les mesures non paramétriques, mais qui peut être important lorsque l'on cherche (comme cela est fait dans la partie 3) à déterminer la fonction mathématique de la distribution.

3. Enfin, il évite les cas extrêmes liés à l'utilisation d'un indicateur de taille obsolète. L'identification des cas extrêmes (réalisée dans la partie II) a en effet montré que les cas les plus extrêmes correspondent à des entreprises qui réalisent des investissements importants durant l'année (l'actif total augmente considérablement cette année-là) tout en générant une forte perte. Cela peut s'expliquer économiquement soit par un mauvais choix d'investissement, soit par un coût lié à

l'entrée sur le nouveau marché. Autrement dit, la perte de l'année N est importante, mais plus ou moins proportionnelle à l'investissement réalisé dans l'année. Mais comparée au montant de l'actif de l'année précédente, le ratio devient extrême. Ce dernier point, qui apparaît dans 11 cas sur les 22 identifiés, est inhérent à la nature même de l'indicateur $AT_{(n-1)}$ qui met en relation deux variables issues de deux exercices comptables différents, et disparaît avec l'utilisation de (AT-IB).

a.2. Le chiffre d'affaires

Comme cela a déjà été observé dans la section C du chapitre IV, le chiffre d'affaires utilisé comme variable de mise à l'échelle n'implique aucune transformation[117]. Par ailleurs, la base n'a pas été complétée (comme cela avait été le cas dans la partie II pour la variable AT), mais certaines valeurs suspectes ont cependant été examinées. Le détail de ces valeurs figure en Annexe 7 : Détail de l'examen de la variable CA dans la base.

Il apparaît ainsi un certain nombre de chiffres d'affaires négatifs ou nuls. Ces 27 observations[118] ont été supprimées. Les variations extrêmes (variation positive supérieure à 10 fois le chiffre d'affaires de l'année précédente ou variation négative inférieure à 90 %) ont été identifiées. Certaines de ces variations représentent manifestement des phases de transition. Le nouveau niveau d'activité est confirmé les années suivantes. Ces cas ont donc été conservés. Ont également été conservées les fortes baisses l'année de sortie de cote. Neuf cas ont été examinés de plus près, et quatre ont été corrigés ou supprimés. Il est utile de souligner que une valeur figurait dans la base pour 1 898,820 au lieu de 1 898 820. Cette correction est importante car un dénominateur extrême conduit le ratio à se situer proche du seuil étudié. En définitive, le nombre d'observations après correction est le suivant (Tableau 47) :

[117] On n'utilise ni $CA_{(n-1)}$, ni (CA-IB).

[118] On remarque que ce sont le plus souvent des entrées dans la cote.

Tableau 47 : Observation de la variable CA dans la base après correction, et comparaison avec la variable IB

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
IB	291	296	376	423	617	702	754	798	791	734	698	659	611	7 750
CA	288	291	374	419	613	689	738	773	773	674	690	650	598	7 570
CA mais pas IB	0	0	1	1	1	1	1	1	1	1	2	1	1	12
IB mais pas CA	3	5	3	5	5	15	18	26	19	61	10	10	14	194
IB et CA	288	291	373	418	612	687	736	772	772	673	688	649	597	7 556

La base est donc constituée de 7 556 observations, soit 97,5 % du total des 7 750 observations pour lesquelles le résultat (IB) est disponible. Étudier la variable « résultat mis à l'échelle par le chiffre d'affaires » conduit donc à la perte de 2,5 % des données seulement.

a.3. La valeur de marché

La mise à l'échelle par la valeur de marché est la plus souvent utilisée par les auteurs anglo-saxons (Burgstahler et Dichev, 1997 ; Kasznik et McNichols, 2002 ; Kinney, Burgstahler et Martin, 2002 ; Matsumoto, 2002 ; Burgstahler et Eames, 2003 ; Burgstahler, Elliott et Hanlon, 2003 ; Dechow, Richardson et Tuna, 2003 ; Holland et Ramsey, 2003 ; Phillips, Pincus et Rego, 2003 ; Payne et Thomas, 2004 ; Xue, 2004 ; Brown et Caylor, 2005 ; Shen et Chih, 2005 ; Burgstahler et Eames, 2006 ; Beaver, McNichols et Nelson, 2007). Ils utilisent la valeur de début d'année ($VM_{(n-1)}$). L'étude reprend cette même variable.

L'effet de l'utilisation de la variable observée l'année précédente a déjà été décrit dans le chapitre IV ; elle conduit à perdre les données de la première année ainsi que les années d'entrée dans la cote. La base comporte 6 689 observations de la variable VM, dont 253 sans le résultat correspondant. Il y a donc 6 436 observations potentielles. Mais après mise à l'échelle, la base ne comporte plus que 5 500 observations de la variable $IB/VM_{(-1)}$, soit une perte de 936 observations pour des causes de décalage (soit 14,5 %).

Tableau 48 : Observation de la variable VM dans la base, et comparaison avec la variable IB

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
IB	291	296	376	423	617	702	754	798	791	734	698	659	611	**7750**
VM														6689
VM (-1)	0	160	203	291	367	514	556	580	603	735	712	698	651	6070
VM(-1) mais pas IB	0	0	0	0	11	34	62	85	72	85	79	73	69	570
IB mais pas VM (-1)	291	136	173	132	261	222	260	303	260	84	65	34	29	2250
IB et VM	0	160	203	291	356	480	494	495	531	650	633	625	582	**5500**
														70.97%

Une recherche des cas étranges a été menée, mais il est apparu qu'il n'y avait aucune valeur négative ou strictement nulle. Un cas extrême a été identifié : le maximum en 1995 s'élève à 237 135 alors que le maximum de toutes les autres années est 134 532. Cette valeur est assez exceptionnelle. Elle appartient à la société Elf Antar Gaz (différente de Elf aquitaine) présente dans la base de 1995 à 1997, mais dont on ne possède la valeur de marché qu'en 1995. Le ratio IB_{1996}/VM_{1995} = 1 058 / 237 135 = 0,0044 est considéré comme un résultat faiblement bénéficiaire. Par comparaison, AT_{1995} = 37 101 et CA_{1995} = 31 639. Seule la mise à l'échelle par la valeur de marché conduit à considérer cette valeur près du seuil du résultat nul. Cela ne justifie pas pour autant la suppression de cette valeur.

Aucune correction de la base n'a donc été effectuée. En définitive, l'étude de la variable résultat mise à l'échelle par $VM_{(-1)}$ se fait sur 5 500 observations, soit 71 % des observations par rapport à l'étude de la seule variable résultat. C'est la variable pour laquelle le nombre de données manquantes est le plus important.

b. Analyse réduite aux observations communes

Dans cette première approche, seul le sous-échantillon composé des entreprises dont toutes les informations sont disponibles est retenu.

Tableau 49 : Nombre d'observations communes aux trois dénominateurs

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total	
Nombre de résultats (IB)	291	296	376	423	617	702	754	798	791	734	698	659	611	7750	**Soit**
IB et AT et CA et VM	0	159	203	290	355	479	491	491	527	616	624	618	573	5426	**70.01%**

Les 5 426 données sur 13 ans[119] sont ensuite agrégées et les caractéristiques des distributions[120] sont observées. Les deux combinaisons issues de l'actif total sont étudiées pour contrôler la neutralité de ces deux transformations.

Tableau 50 : Caractéristiques des distributions des variables de mise à l'échelle

	CA93a04a	VM92a03a[121]	AT93a04a	ATmIB~04
N	5426	5426	5426	5426
max	122 700	237 136	905 938	901 270
min	0.035	0.172	1.422	0.638
mean	2 472.83	1 616.56	7 991.62	7 934.87
p50	175.69	120.58	245.38	236.96
sd	8 034	7 282	45 707	45 552
skewness	6.92	13.23	11.54	11.53
kurtosis	69.37	284.17	165.10	164.70

Il apparaît que les quatre distributions ont la même allure : pointues (kurtosis très supérieur à 3[122]) et asymétrique vers la droite (skewness nettement supérieur à zéro). Cela s'explique dans la mesure où il y a relativement peu de grandes entreprises, et relativement beaucoup de petites entreprises.

b.1. Étude des corrélations

Le Tableau 51 montre que les trois variables de mise à l'échelle sont toutes très significativement corrélées. Le choix du dénominateur ne remet donc pas en cause l'analyse qui peut être faite des distributions de résultat. L'utilisation de chacune de ces trois variables de mise à l'échelle doit conduire à des conclusions similaires.

Par ailleurs, la corrélation est proche de 1 entre le total actif annuel, et le total actif diminué du résultat. Cette transformation paraît donc neutre pour l'analyse.

[119] Fonction sous stata : stack AT93 CA92 VB93 AT94 CA93 VB94 AT95 CA94 VB95 AT96 CA95 VB96 AT97 CA96 VB97 AT98 CA97 VB98 AT99 CA98 VB99 AT00 CA99 VB00 AT01 CA00 VB01 AT02 CA01 VB02 AT03 CA02 VB03 AT04 CA03 VB04, into(AT93a04a CA92a03a VB93a04a)

[120] Fonction sous stata : tabstat AT93a04a CA93a04a VB92a03a ATmIBde93a04, statistics(count max min mean median sd skewness kurtosis) columns(variables)

[121] La variable AT est collectée de 1992 à 2003 puisque il y a décalage d'un an. Les autres variables sont collectées de 1993 à 2004.

[122] Le kurtosis nul correspond à une distribution plate. Le kurtosis de 3 correspond à la loi normale.

Tableau 51 : Corrélations des variables de mise à l'échelle entre elles

Corrélations[123]	VM92a03a	CA93a04a	AT93a04a	ATmIB~04
VM92a03a	1			
CA93a04a	0.6926	1		
P-value	0.000			
AT93a04a	0.376	0.6209	1	
P-value	0.000	0.000		
ATmIBde93a04a	0.3745	0.6187	0.9999	1
P-value	0.000	0.000	0.000	

Pour affiner l'analyse, la même étude des corrélations a été menée sur les seules entreprises proches du seuil du résultat nul.

b.2. Étude des corrélations sur les seules entreprises proches du seuil

Sur Excel, les ratios IB_n/CA_n, $IB_n/VM_{(n-1)}$, et $IB_n/(AT_n-IB_n)$ ont été calculés uniquement sur la base réduite aux observations communes. Puis parmi ces données, seules les observations dont l'un au moins des trois ratios pour chaque entreprise/année est compris dans l'intervalle]-0,005 ; 0,005[ont été sélectionnés. Le nombre d'observations proches du seuil du résultat nul sont alors les suivantes :

Tableau 52 : Nombre d'entreprises faiblement déficitaire ou faiblement bénéficiaire selon un au moins des ratios

Année	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Nombre d'observations proches du seuil selon un au moins des trois critères	47	54	63	62	80	100	73	97	111	110	97	71	965

Les données ont alors été transférées sous Stata pour créer trois variables agrégées[124]. Le Tableau 53 indique les caractéristiques des trois variables ainsi obtenues[125].

[123] Fonction sous Stata : pwcorr VB92a03a CA93a04a ATmIBde93a04, sig

[124] Fonction sous Stata : stack CA93 VB92 ATmIB93 CA94 VB93 ATmIB94 CA95 VB94 ATmIB95 CA96 VB95 ATmIB96 CA97 VB96 ATmIB97 CA98 VB97 ATmIB98 CA99 VB98 ATmIB99 CA00 VB99 ATmIB00 CA01 VB00 ATmIB01 CA02 VB01 ATmIB02 CA03 VB02 ATmIB03 CA04 VB03 ATmIB04, into(CA93a04b VB92a03b ATmIB93a04b)

[125] Fonction sous Stata : tabstat CA93a04b VB92a03b ATmIB93a04b, statistics(count max min mean sd skewness median kurtosis) columns(variables).

Tableau 53 : Comparaison des dénominateurs des entreprises faiblement déficitaires ou faiblement bénéficiaires

stats	CA93a04b	VM92a03b	ATml~04b
N	965	965	965
max	98884	237136.7	901270
min	0.1282022	0.2673416	3.11011
mean	4862.816	2100.131	31461.64
P50	420.7228	132.7073	1070.33
sd	11232.93	9589.076	101528.7
skewness	4.378158	16.63961	5.071925
kurtosis	28.17629	381.3742	32.39747

Les corrélations sont présentées dans la Tableau 54[126] :

Tableau 54 : Corrélations des dénominateurs des entreprises identifiées comme faiblement bénéficiaires/déficitaires

	VM92a03b	CA93a04b	ATml~04b
VM92a03b	1		
CA93a04b	0.5586	1	
P-value	0.000		
ATmlB93a04b	0.4803	0.8102	1
P-value	0.000	0.000	

Il apparaît à nouveau une corrélation très significative. Les coefficients sont un peu plus élevés pour les valeurs proches du seuil du zéro. En conclusion, changer de dénominateur ne doit pas avoir de forte influence sur les résultats des études sur les seuils.

Il n'en demeure pas moins que ces calculs ont été réalisés à partir d'une base ajustée pour que toutes les données soient disponibles. Or le dénominateur VM est moins souvent disponible et que une année d'observation est perdue. Autrement dit, l'étude des corrélations est effectuée sur 70 % des observations. Pour mesurer l'impact de ce biais de sélection, l'étude des dénominateurs est approfondie en comparant les mesures d'irrégularités comme si les trois dénominateurs étaient utilisés indépendamment les uns des autres.

[126] Fonction sous Stata : pwcorr VB92a03b CA93a04b ATmlB93a04b, sig

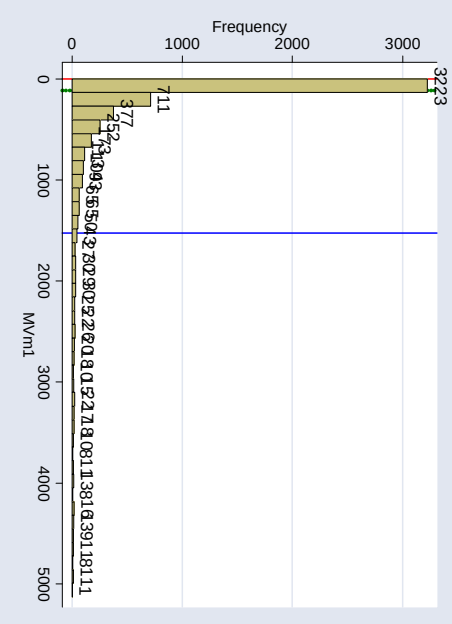
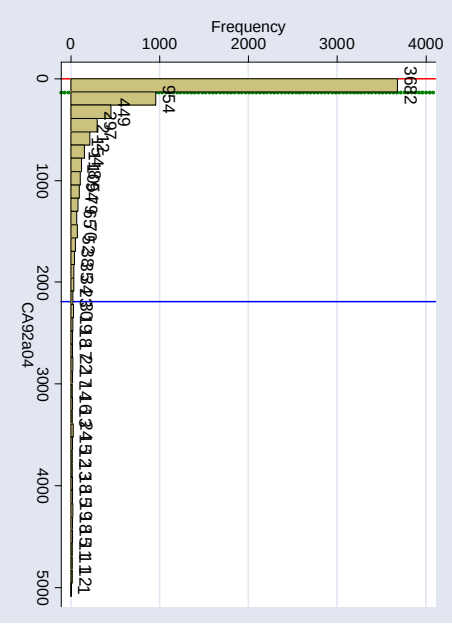
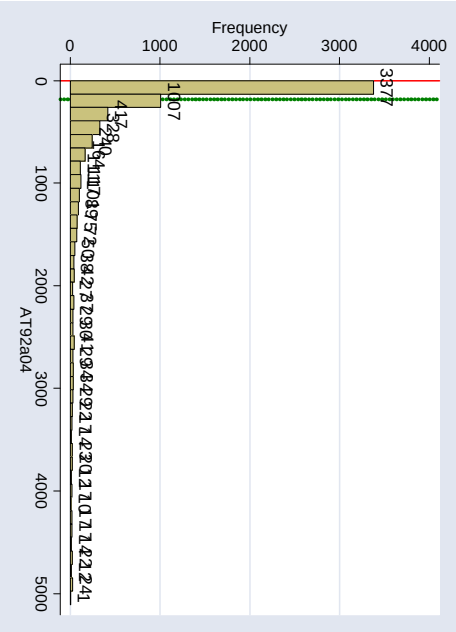
c. Analyse étendue à toutes les observations disponibles

Selon le choix du dénominateur retenu, la base sera différente du fait des valeurs manquantes. Il est donc intéressant de comparer les distributions des dénominateurs en retenant l'ensemble des observations disponibles pour chaque dénominateur (et non pas comme précédemment en ne prenant que les observations communes) afin d'évaluer l'impact concret de ces biais de sélection sur les mesures. Au préalable, une réflexion est menée sur la mesure d'une « petite performance » dont la valeur peut être différente selon le dénominateur utilisé (sous partie c.1). La deuxième sous-partie (c.2) compare les entreprises identifiées comme proches du seuil selon les différents dénominateurs. La dernière sous-partie (c.3) compare les irrégularités mesurées selon les différents dénominateurs.

c.1. Comparaison des trois distributions et notion de petite performance

La Figure 82 représente les distributions des trois variables (agrégées sur 13 ans) de mise à l'échelle.

Figure 82 : Distribution des trois variables de mise à l'échelle

Valeur de marché	Chiffre d'Affaires	Actif Total	
			
6 070	7 568	7 762	Nombre
237 136	122 700	905 938	Max
0.03097	0.032	0.037	Min
1 529.1	2 191.083	7 419.43	Moyenne
116.8	139.2	184.6	Médiane
6 987.8	7 311.7	42368.7	Écart type
301.3	75.3	169.3	Kurtosis
13.6	7.1	11.5	Skewness
histogram MVm1 if MVm1<5000, width(100) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts (mlabsize(medium)) xline(0, lcolor(red)) xline(116,lcolor(green) lwidth(thick) lpattern(dot)) xline(1529,lcolor(blue)) yscale(range(. .4000))	histogram CA92a04 if CA92a04<5000, width(100) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts (mlabsize(medium)) xline(0, lcolor(red)) xline(139,lcolor(green) lwidth(thick) lpattern(dot)) xline(2191,lcolor(blue)) yscale(range(. .4000))	histogram AT92a04 if AT92a04<5000, width(100) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts (mlabsize(medium)) xline(0, lcolor(red)) xline(184,lcolor(green) lwidth(thick) lpattern(dot)) xline(7419,lcolor(blue)) yscale(range(. .4000))	Fonction (sous stata)

Observation des trois distributions

Il apparaît une allure similaire dans les trois cas, caractérisée par un mode situé juste au dessus du zéro, et une pente décroissante non linéaire. La répartition de la taille des entreprises n'est pas homogène. Il y a un très grand nombre de très petites entreprises, et un très petit nombre de très grandes entreprises, et ce, quelle que soit la variable choisie.

Le Kurtosis permet de classer les distributions de la plus concentrée à la moins concentrée. Il apparaît que la valeur de marché est plus pointue que l'actif total et enfin que le chiffre d'affaires. Les trois médianes offrent des ordres de grandeur similaires, mais les moyennes et les maximums peuvent présenter des écarts très importants. La distribution de l'actif total est celle qui présente les plus grandes valeurs extrêmes.

La nombreuse présence des sociétés de petite taille explique la forte concentration de la variable résultat (non mise à l'échelle) autour du zéro.

Notion de petite performance

Pour effectuer la comparaison des entreprises identifiées comme faiblement bénéficiaires ou déficitaires, il faut établir une norme permettant de considérer qu'un résultat est « petit ». Autrement dit, il faut fixer une largeur d'intervalle. Pour ce faire, il est possible de retenir, soit une largeur commune aux trois distributions, soit trois largeurs différentes.

Comme les caractéristiques des variables de mise à l'échelle sont différentes, il faut adapter la largeur à chaque dénominateur. Pour déterminer la largeur d'intervalle, il faut, soit utiliser la formule théorique, soit fixer arbitrairement un seuil. La largeur théorique a été calculée pour chaque année selon chaque dénominateur (voir Annexe 8 : Largeurs théoriques en fonction du dénominateur). Les valeurs obtenues oscillent d'une année sur l'autre entre 0,01 et 0,04. Mais ce calcul de largeur est construit pour optimiser l'observation d'une distribution. Il dépend de l'intervalle interquartile et du nombre d'observations. Or, ces deux indicateurs, et notamment le nombre d'observations, n'ont aucun rapport avec la notion de « petite performance ». C'est pourquoi cette approche n'est pas retenue.

Dans la littérature, la largeur d'intervalle retenue est souvent de 0,005. C'est la largeur retenue dans la section A. Cette section étudie le résultat réduit par l'actif total. Si la notion de « petite performance » est définie comme une performance inférieure à 0,005 fois (actif total - résultat net), puisque sur 13 ans l'actif total médian est de 185, la petite performance sera définie comme étant inférieure à 185 x 0,005 = 0,9229. Autrement dit,

les entreprises effectuent des petites pertes ou des petits gains si elles sont situées dans l'intervalle] -0,9229 ; 0,9229 [.

La taille médiane exprimée en chiffre d'affaires est de 139 et de 116 en valeur de marché. Il en découle que le bénéfice de 0,9229 considéré comme petit pour l'entreprise médiane représente 0,00664[127] CA et 0,00796 VM. En conclusion, si l'on veut comparer des choses comparables, il faut utiliser une largeur de 0,005 avec la variable AT, 0,007 avec la variable CA et 0,008 avec la variable VM (aux arrondis près). Les entreprises faiblement bénéficiaires et faiblement déficitaires ont été identifiées (Tableau 55) dans ces trois situations.

Tableau 55 : Identification des entreprises faiblement bénéficiaires ou déficitaires

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
IB/ AT-IB	**largeur de 0,005**													
nombre	46	43	51	54	59	68	55	43	44	55	53	43	29	643
En %														8.30%[128]
IB/ CA	**largeur de 0,007**													
nombre	33	25	32	39	42	39	48	34	39	46	51	54	39	521
En %	CA													6.90%[129]
IB/ VM	**largeur de 0,008**													
nombre	0	7	3	14	5	12	14	9	24	43	30	23	17	201
En %														3.65%[130]

On remarque que, malgré les efforts déployés afin de définir une notion de « petite performance » relative au dénominateur, la part des entreprises situées sur les intervalles « petites pertes » et « petits gains » dans la base est inférieure lorsqu'on utilise une mise à l'échelle par la valeur de marché, et dans une moindre mesure par le chiffre d'affaires.

Tableau 56 : Concentration des distributions

	RsurA~04	RsurC~04	RsurV~04
N	7 746	7 556	5 500
kurtosis	823.37	1 207.20	1 157.76

Les distributions en mettant à l'échelle par le chiffre d'affaires comme par la valeur de marché sont plus concentrées (Kurtosis élevé) que la distribution en mettant à l'échelle par le total actif. Le seuil du résultat nul ne se situe pas en général sur le sommet de la

[127] 0,9229/139 = 0,00664 et 0,9229/116 = 0,00796.

[128] 8,30% des 7746 observations de IB/(AT-IB)

[129] 6,92% des 7556 observations de IB/(CA)

[130] 3,65% des 5500 observations de IB/VM(-1)

distribution. Cette raison peut en partie expliquer que les effectifs qui encadrent ce seuil sont plus importants lorsque l'on met à l'échelle par le total actif (8,3 % contre 6,9 % et 3,7 %).

Une base réduite aux entreprises réalisant de petites performances (proches de zéro) selon une au moins des formules a ensuite été créée.

c.2. Comparaison des entreprises identifiées selon les trois variables

Dans quelle mesure les entreprises identifiées par une méthode comme étant proche du seuil sont-elles également identifiées par les autres ? Peut-on s'attendre à une différence significative en fonction du dénominateur ? Pour répondre à ces questions, les observations identifiées comme étant proches du seuil résultat nul (donc entreprises réalisant des petits gains ou des petites pertes) par au moins un des trois dénominateurs sont comparées.

Tableau 57 : Entreprises faiblement bénéficiaires ou déficitaires

	petites perf selon R/(AT-R)	petites perf selon R/CA	petites perf selon R/VM-1
Définition de la petite performance	Abs(R/(AT-R))<0.005	Abs(R/CA)<0.007	Abs(R/VM-1)<0.008
Nombre de cas identifiés	643	521	201

Dans un premier temps, les entreprises/année identifiées comme proches du seuil résultat nul en utilisant l'actif total et le chiffre d'affaires comme variable de mise à l'échelle sont comparées. Puis elles sont comparées aux mesures effectuées avec la valeur de marché.

c.2.1. Comparaison des cas identifiés en utilisant les variables CA et AT

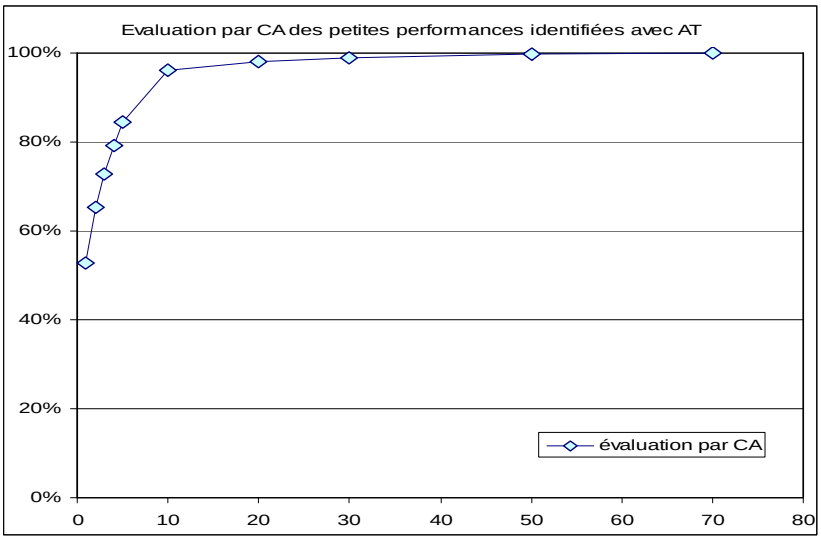
Parmi les 643 observations de petites performances identifiées en utilisant l'actif total comme dénominateur, 18 correspondent à des chiffre d'affaires manquants. L'étude porte donc sur les 625 petites performances identifiées en mettant à l'échelle par l'actif total et pour lesquelles il est possible de comparer le ratio IB/(AT-IB) au ratio IB/CA.

330 observations sont identifiées par les deux dénominateurs comme étant des petites performances, soit près d'une observation sur deux. En tenant compte des observations

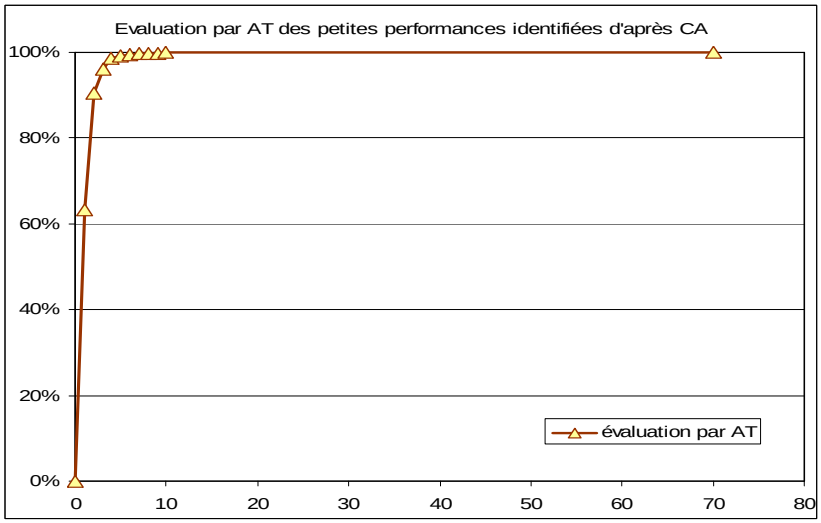
situées sur l'intervalle adjacent, les observations communes s'élèvent à 408 (soit 65,3 %). 601 observations (soit 96,2 %) sont éloignées de moins de 10 intervalles.

La Figure 83 fait apparaître que les « très » petites performances identifiées en utilisant l'actif total comme dénominateur ne sont pas toujours identifiées comme telles en utilisant le chiffre d'affaires. Ceci étant dit, moins de 4 % des observations faites en utilisant l'actif total dépassent de 10 fois le montant de la « petite performance » définie pour la mise à l'échelle par le chiffre d'affaires. Autrement dit, pour le plus grand nombre d'observations, même si des écarts de mesure sont sensibles, il n'y a pas de profonde différence. Une très petite performance identifiée en utilisant l'actif total sera dans 96 % des cas identifiée comme étant une petite performance en utilisant le chiffre d'affaires au dénominateur.

Figure 83 : Comparaison des entreprises identifiées comme faiblement bénéficiaires/déficitaires en utilisant les variables CA et AT



Les 643 petites performances identifiées en utilisant AT au dénominateur ne sont pas toutes confirmées en utilisant CA.



Les 521 petites performances identifiées en utilisant CA au dénominateur sont toutes identifiées en utilisant AT.

En renversant l'analyse, le second graphe de la Figure 83 s'intéresse aux 521 petites performances identifiées en utilisant le chiffre d'affaires comme dénominateur. Dans tous les cas, l'actif total correspondant est disponible. Si 330 observations (63,3 %) sont identifiées par les deux dénominateurs comme étant des petites performances, les autres cas identifiés en mettant à l'échelle par le chiffre d'affaires sont situés à moins de 10 fois la petite performance déterminée en utilisant l'actif total comme dénominateur. 471 (plus de 90 %) sont situés à moins d'un intervalle de distance. Autrement dit, toutes les petites performances identifiées en utilisant le chiffre d'affaires au dénominateur sont identifiées comme étant également des petites performances en utilisant l'actif total, alors que l'inverse n'est pas toujours vrai.

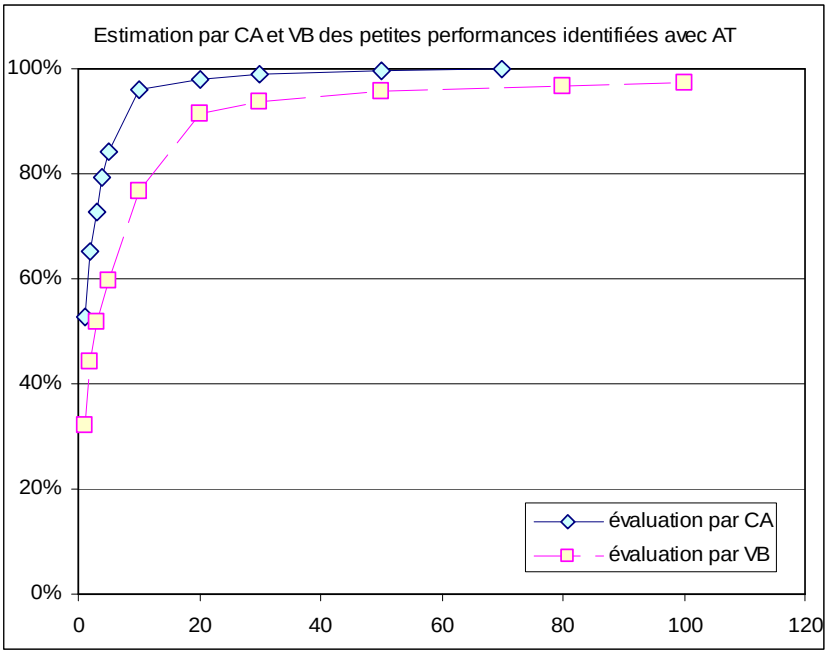
En conclusion, mettre à l'échelle par l'actif total ou par le chiffre d'affaires semble relativement indifférent (au problème de la largeur d'intervalle près).

c.2.2. Comparaison avec les cas identifiés en utilisant la valeur de marché

Le pourcentage des entreprises faiblement performantes en mettant l'actif total au dénominateur, identifiées également en mettant la valeur de marché au dénominateur, est représenté par les carrés de la Figure 84. Parmi les 643 petites performances identifiées en utilisant l'actif total comme dénominateur, 212 correspondent à des valeurs boursières manquantes. Sur les 431 petites performances restantes, seules 138 observations sont identifiées par les deux dénominateurs comme étant des très petites performances (soit près de une observation sur trois). Si l'on étend l'observation à l'intervalle adjacent, seules 191 observations (44,3 %) sont communes. 330 observations (soit 76,6 % seulement) sont éloignées de 10 intervalles ou moins. Autrement dit, les mesures effectuées en utilisant la valeur de marché au dénominateur n'identifient pas nécessairement les mêmes entreprises comme étant faiblement performantes. Ces différences sont importantes..

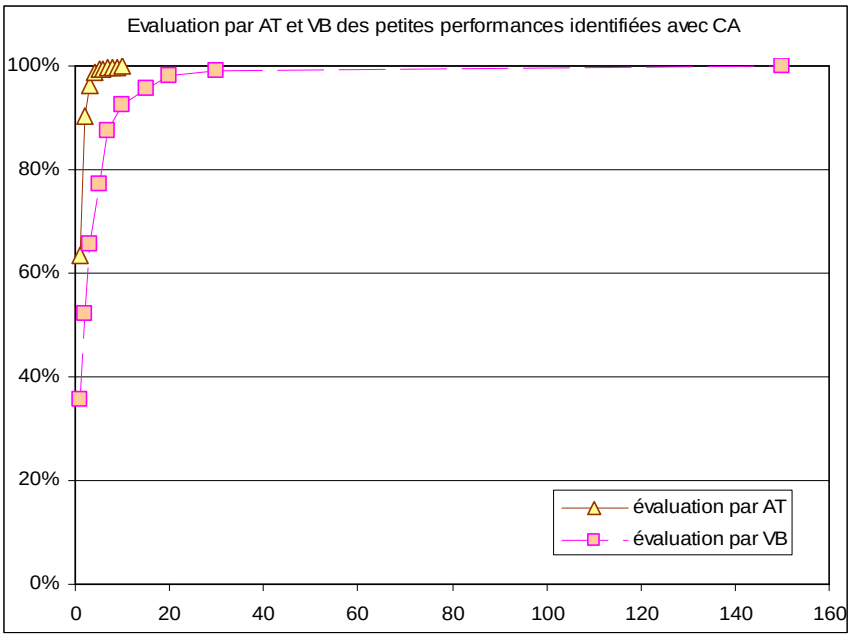
Paradoxalement, il apparaît qu'un petit nombre d'entreprises identifiées comme proche du résultat nul en utilisant l'actif total au dénominateur sont identifiées comme étant très performantes en utilisant la valeur de marché (ratio plus de 100 fois supérieur).

Figure 84 : Identification, en utilisant les variables CA et VM, des petites performances identifiées en utilisant la variable AT



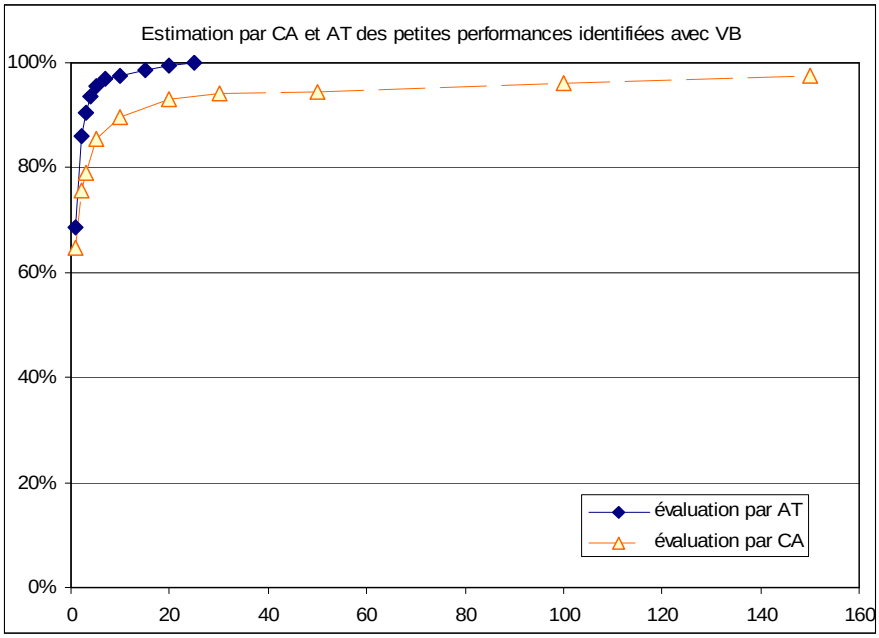
Les petites performances identifiées en utilisant la valeur de marché sont ensuite comparées aux petites performances identifiées en utilisant le chiffre d'affaires, et sont représentées par les carrés de la Figure 85. Sur les 521 petites performances identifiées en utilisant le chiffre d'affaires, 155 ne peuvent être comparées à la performance mise à l'échelle par la valeur de marché car les données sont manquantes. Sur les 366 observations restantes, 130 observations (35,52 %) sont identifiées par les deux dénominateurs comme étant des petites performances. Les performances mises à l'échelle par la valeur de marché sont une fois de plus sensiblement différentes. Même si plus de 90 % des petites performances identifiées en utilisant chiffre d'affaires ne sont pas éloignées de plus de 10 intervalles des petites performances identifiées par valeur de marché, un petit nombre de petites performances identifiées en utilisant la variable CA correspondent à des performances importantes lorsqu'elles sont mises à l'échelle par la valeur de marché.

Figure 85 : Identification, en utilisant les variables AT et VM, des petites performances identifiées en utilisant la variable CA



L'actif total et le chiffre d'affaires sont toujours disponibles pour les 201 observations identifiées comme étant des très petites performances avec la valeur de marché au dénominateur. La Figure 86 montre que l'utilisation de la variable AT permet d'identifier plus de 90 % de ces entreprises à moins de deux intervalles de distance de l'intervalle « très petite performance » (losanges bleus). En revanche, il apparaît que la mise à l'échelle par le chiffre d'affaires (triangles jaunes) peut conduire à considérer comme très performantes des entreprises dont le résultat relatif est considéré comme quasi nul en mettant à l'échelle par la valeur de marché.

Figure 86 : Identification, en utilisant les variables AT et CA, des petites performances identifiées en utilisant la variable VM



En conclusion, il apparaît que la mise à l'échelle par la valeur de marché fournit des mesures relativement atypiques par rapport aux deux autres variables. Les mises à l'échelle par les variables AT et CA sont en revanche convergentes. Il apparaît que la mise à l'échelle par la variable AT permet d'identifier la plupart des entreprises considérées comme peu performantes par la variable VM. Par ailleurs, la base Global Vantage est beaucoup moins complète en ce qui concerne la valeur de marché. En conclusion, l'utilisation de la variable AT semble préférable.

c.3. Mesure et comparaison des irrégularités

Méthode

Les irrégularités des effectifs situés sur l'intervalle (1) juste supérieur, et (2) juste inférieur au seuil du résultat nul, sont mesurées. La largeur d'intervalle utilisée est de 0,005 pour IB/(AT-IB), 0,07 pour IB/CA et 0,008 pour IB/VM_{-1}. L'effectif attendu est estimé par interpolation linéaire sur les 8 intervalles qui entourent les deux effectifs étudiés (4 inférieurs et 4 supérieurs). L'irrégularité est exprimée en pourcentage de l'effectif attendu.

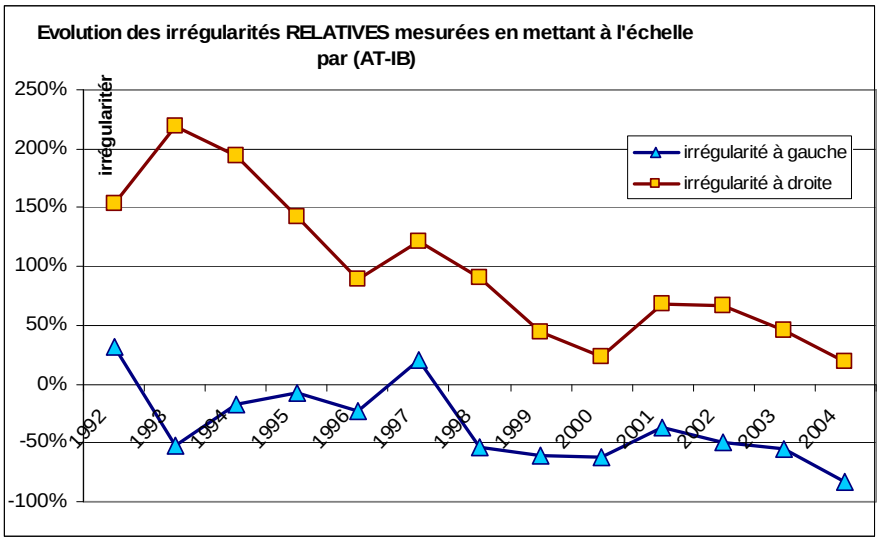
Irrégularités mesurées en mettant à l'échelle par l'actif total

Les calculs ont été effectués dans la section A. Les résultats sont rappelés dans le Tableau 58 ci-dessous :

Tableau 58 : Irrégularités annuelles mesurées en mettant à l'échelle par l'actif total

irrégularité relative de IB/(AT-IB)		
	gauche	Droite
1992	31.5%	153.9%
1993	-52.8%	219.4%
1994	-17.0%	193.8%
1995	-7.3%	141.6%
1996	-22.7%	88.4%
1997	20.3%	120.7%
1998	-53.6%	90.8%
1999	-60.5%	44.4%
2000	-61.5%	23.4%
2001	-36.6%	68.1%
2002	-49.0%	66.9%
2003	-54.6%	45.0%
2004	-83.0%	19.0%

Représentation graphique



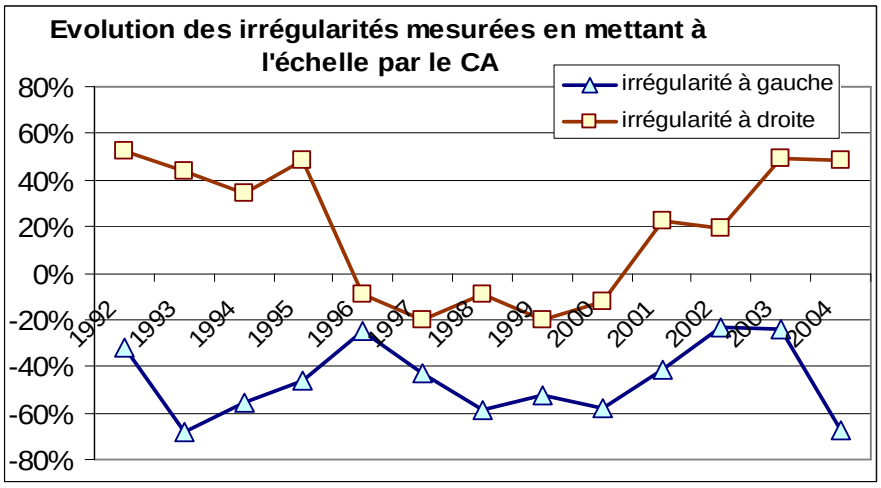
Irrégularités mesurées en mettant à l'échelle par le chiffre d'affaires

Les mesures sont effectuées à partir des régressions dont la mise en œuvre est détaillée en Annexe 9 : Irrégularités mesurées en mettant le résultat à l'échelle par le chiffre d'affaires. Ces treize régressions annuelles permettent d'estimer des effectifs théoriques à droite comme à gauche du seuil, et à mesurer les irrégularités qui sont récapitulées et représentées graphiquement dans le tableau ci-dessous.

Tableau 59 : Irrégularités annuelles mesurées en mettant à l'échelle par le chiffre d'affaires

irrégularité relative de IB/CA	gauche	Droite
1992	-31,9%	52.1%
1993	-68,4%	43.7%
1994	-55,2%	34.3%
1995	-46,4%	48.4%
1996	-24,9%	-9.4%
1997	-43,3%	-20.2%
1998	-58,6%	-8.9%
1999	-52,4%	-20.4%
2000	-57,6%	-12.0%
2001	-41,1%	22.6%
2002	-22,9%	19.4%
2003	-23,9%	49.1%
2004	-67,1%	48.2%

Représentation graphique



On remarque que l'irrégularité à gauche du seuil du résultat nul est toujours négative conformément à la théorie et aux travaux antérieurs. En revanche, l'évolution de l'irrégularité à droite est plus chaotique. Si elle est en général positive conformément aux attentes, elle est (très légèrement) négative entre 1996 et 2000.

Irrégularités mesurées en mettant à l'échelle par la valeur de marché

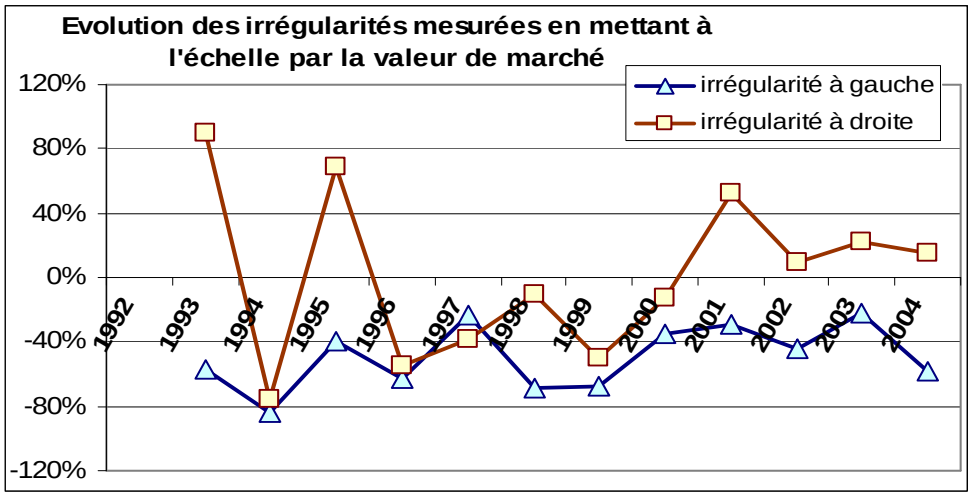
Les mesures sont effectuées à partir des régressions dont la mise en œuvre est détaillée en Annexe 10 : Irrégularités mesurées en mettant le résultat à l'échelle par la valeur de marché. Ces treize régressions annuelles permettent d'estimer des effectifs théoriques à droite comme à gauche du seuil, et à mesurer les irrégularités qui sont récapitulées dans le tableau ci-dessous.

Tableau 60 : Calcul des irrégularités (résultat mis à l'échelle par la variable VM)

	Intervalle à gauche du seuil				Intervalle à droite du seuil			
	Observation	Estimation	écart	irrégularité (en %)	Observation	Estimation	écart	irrégularité (en %)
1992	-	-	-	-	-	-	-	-
1993	1	2.33	-1.33	-57.1%	6	3.17	2.83	89.2%
1994	1	6.36	-5.36	-84.3%	2	8.39	-6.39	-76.2%
1995	3	4.96	-1.96	-39.6%	11	6.54	4.46	68.3%
1996	2	5.41	-3.41	-63.1%	3	6.59	-3.59	-54.4%
1997	6	7.81	-1.81	-23.2%	6	9.69	-3.69	-38.1%
1998	3	9.71	-6.71	-69.1%	11	12.29	-1.29	-10.5%
1999	3	9.37	-6.37	-68.0%	6	12.13	-6.13	-50.6%
2000	9	13.83	-4.83	-34.9%	15	17.17	-2.17	-12.6%
2001	12	16.96	-4.96	-29.2%	31	20.29	10.71	52.8%
2002	9	16.13	-7.13	-44.2%	21	19.12	1.88	9.8%
2003	8	10.22	-2.22	-21.7%	15	12.28	2.72	22.1%
2004	4	9.50	-5.50	-57.9%	13	11.25	1.75	15.5%

Ces irrégularités sont représentées graphiquement ci-dessous :

Figure 87 : Représentation graphique des irrégularités ; résultat mis à l'échelle par valeur de marché

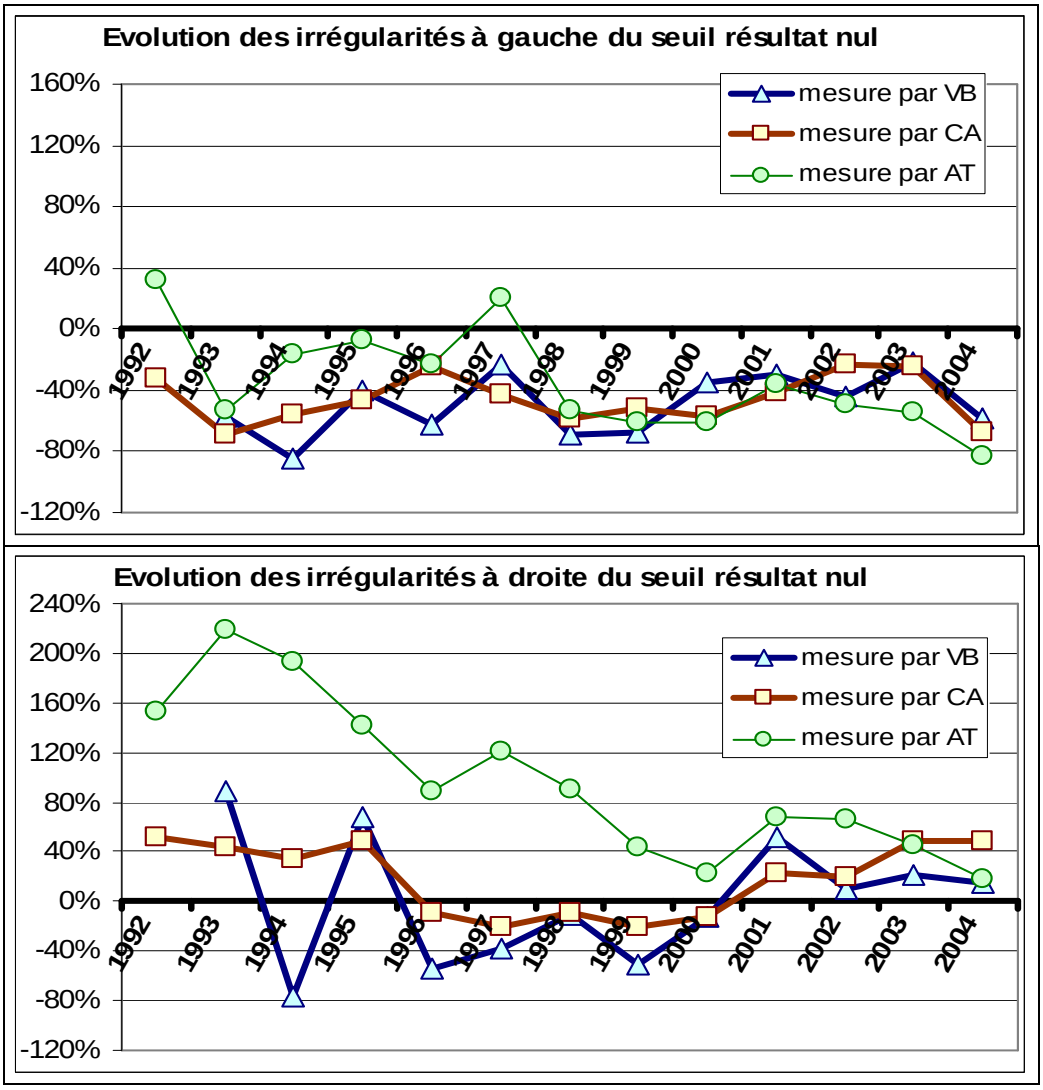


On remarque que l'irrégularité à gauche du seuil du résultat nul est toujours négative conformément à la théorie et aux travaux antérieurs. En revanche, l'évolution de l'irrégularité à droite est plus chaotique. Elle est (faiblement) négative 6 fois sur 12.

Comparaison des mesures effectuées avec les trois dénominateurs

Le principal résultat de l'étude est que les irrégularités à droite et à gauche du seuil sont observées dans presque tous les cas. Autrement dit, le dénominateur ne peut à lui seul expliquer la discontinuité. La présomption de gestion du résultat en est confortée.

Figure 88 : Comparaison des irrégularités mesurées en mettant à l'échelle par les différents dénominateurs



Cependant, il apparaît que le choix de la variable peut entraîner des différences importantes dans la mesure des irrégularités, et ce, essentiellement en ce qui concerne l'irrégularité à droite du seuil. Les dénominateurs CA et plus encore VM tendent à sous-estimer le nombre d'entreprises dépassant de peu le seuil par rapport aux mesures effectuées avec la variable AT au dénominateur. Ces différences viennent s'ajouter aux limites liées à la largeur des intervalles d'observation et à la méthode utilisée pour estimer l'effectif théorique.

Une explication à la sous-estimation de l'irrégularité en utilisant le chiffre d'affaires et la valeur de marché

La distribution des résultats des entreprises dont le chiffre d'affaires est manquant est comparée à la distribution de l'ensemble de la population. Le Tableau 61 synthétise ces comparaisons aux alentours du seuil résultat nul.

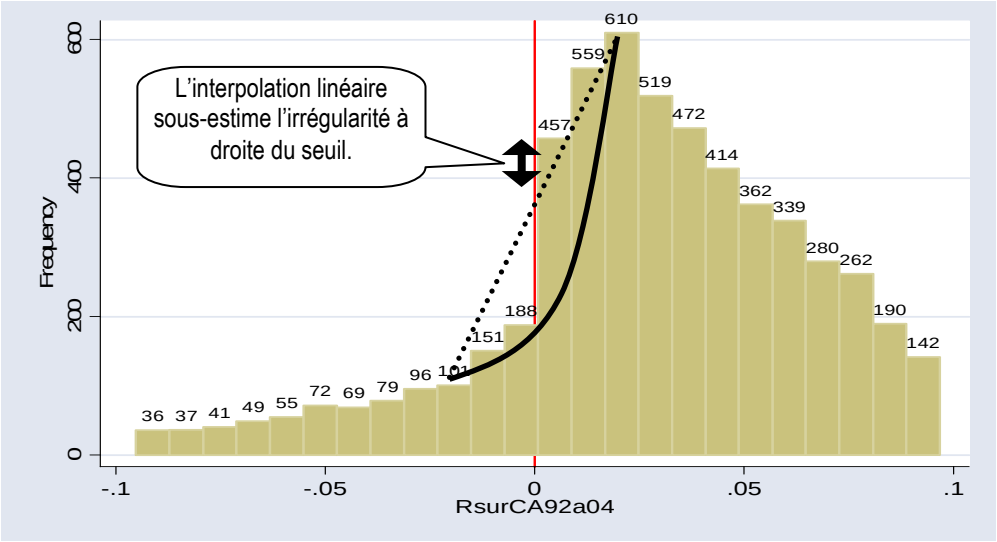
Tableau 61 : Distribution des entreprises dont le chiffre d'affaires est manquant

résultat	[-20;-15[	[-15 ;-10[	[-10 ;-5[	[-5 ;0[	[0 ;5[	[5 ;10[	[10 ;15[	[15 ;20[	total
Toute la population	59	109	217	834	2386	774	456	305	7750
CA manquants	0	1	7	36	82	9	5	2	153
	0.00%	0.92%	**3.23%**	**4.32%**	**3.44%**	1.16%	1.10%	0.66%	**1.97%**

Il apparaît que les entreprises dont le chiffre d'affaires est manquant sont concentrées essentiellement autour du seuil, donc sur la zone d'intérêt. Il y a donc mécaniquement une sous-estimation des effectifs proches du seuil lorsque l'on met à l'échelle par le chiffre d'affaires. Autrement dit, si le nombre d'entreprises légèrement déficitaires est sous-estimé, cela tend à accroître l'irrégularité négative en dessous du seuil. À l'inverse, un nombre d'entreprises faiblement bénéficiaires sous-estimé tend à sous-estimer l'irrégularité à droite du seuil. Cela explique en partie la différence entre les mesures qui utilisent la mise à l'échelle par l'actif total et par le chiffre d'affaires. Il y a un biais de sélection important lorsque le chiffre d'affaires est utilisé.

Par ailleurs, la distribution agrégée sur 13 ans de la variable résultat mise à l'échelle par le chiffre d'affaires est tracée sur la Figure 89. Le seuil du résultat nul étant situé sur une portion fortement croissante de la courbe, il est probable que la méthode d'estimation de l'effectif par interpolation linéaire ne soit pas la mieux adaptée. Elle tend en effet à surestimer l'effectif attendu au dessus du seuil, et donc à sous-estimer l'irrégularité.

Figure 89 : Distribution agrégée sur 13 ans du résultat mis à l'échelle par le chiffre d'affaires (partie centrale de la distribution)



En ce qui concerne l'utilisation de la valeur de marché, il apparaît également (voir Tableau 62) que les entreprises dont le résultat est très proche de zéro (donc situées à proximité du seuil) sont très largement sous-représentées du fait de la concentration en ce point des valeurs manquantes dans la base. Du fait de ce biais de sélection, mettre à l'échelle par la valeur de marché tend à sous-estimer les effectifs observés à droite et à gauche du seuil, donc à surestimer l'écart négatif à gauche, et à sous-estimer le sur-effectif à droite.

Tableau 62 : Distribution des entreprises dont la valeur de marché est manquante

résultat	[-20;-15[	[-15 ;-10[	[-10 ;-5[	[-5 ;0[	[0 ;5[	[5 ;10[	[10 ;15[	[15 ;20[	total
Toute la population	59	109	217	834	2386	774	456	305	7750
VM manquants	4	10	20	158	629	106	40	29	1061
écart	6.78%	9.17%	9.22%	**18.94%**	**26.36%**	13.70%	8.77%	9.51%	**13.69%**

Ces remarques conduisent à penser qu'une méthode non paramétrique de mesure, associée au biais de sélection, explique tout ou partie des écarts de mesure constatés entre les dénominateurs.

3. Conclusion et apports sur l'impact du choix du dénominateur

En conclusion, le choix du dénominateur utilisé pour mettre la variable résultat à l'échelle influence très sensiblement les mesures d'irrégularité. Il n'en demeure pas moins que quelque soit le dénominateur utilisé, une irrégularité est observable autour du seuil du résultat nul. Cette irrégularité est donc imputable à la variable résultat.

Dans l'échantillon, l'irrégularité à gauche est, dans les trois cas, identifiée et mesurée de manière similaire. La mesure de l'irrégularité à droite est plus sensible au choix du dénominateur. L'utilisation de l'actif total comme dénominateur permet de faire apparaître une irrégularité qui n'est pas toujours aussi prononcée en utilisant le chiffre d'affaires ou la valeur de marché. Dans le cadre de cette étude, ces différences sont dues essentiellement à un biais de sélection lié à la composition de la base qui diffère selon le dénominateur étudié. Mais elles sont également renforcées par les limites déjà étudiées dans les sections A et B, à savoir le choix de la largeur des intervalles d'observation, et le choix de la méthode de mesure de l'effectif théorique.

D'après Glaum, Lichtblau et Lindemann (2004), l'utilisation du chiffre d'affaires est préférable dans le cadre des études internationales. Cette affirmation semble devoir être relativisée. Les limites méthodologiques sont trop importantes pour se contenter d'une telle affirmation.

L'étude montre que l'utilisation de la valeur de marché peut s'avérer délicate. Les données concernant cette variable dans la base Global Vantage sont moins complètes. Par ailleurs, le décalage d'un an conduit à la perte d'un grand nombre d'observations, notamment de celles qui entrent ou sortent de la base. Or ces sociétés ont plus de chances de présenter des résultats proches du seuil. De plus, la volatilité de la valeur de marché est à souligner. Elle n'est pas la référence de taille la plus stable pour des comparaisons dans le temps ou dans le cadre d'études internationales. Enfin, comme le soulignent Durtschi et Easton (2005), les marchés n'évaluent pas nécessairement les entreprises déficitaires selon les mêmes critères que les entreprises bénéficiaires. Cela peut influencer significativement la mesure des irrégularités. Plus généralement, la valeur de marché est inadaptée aux études incluant des entreprises non cotées.

Enfin, l'utilisation de l'actif total au dénominateur permet d'identifier les entreprises dont les performances sont proches de zéro avec les autres dénominateurs, alors que le

contraire n'est pas toujours vrai. Pour cette raison, l'utilisation de cette variable est privilégiée. Dans tous les cas, une étude sur les seuils ne peut faire l'économie de contrôler ses résultats en comparant les résultats qu'elle aurait obtenus en utilisant un autre dénominateur. Enfin, elle ne peut faire l'économie d'un examen des valeurs manquantes, dans la base de données, de la variable utilisée pour mettre le résultat à l'échelle, afin notamment d'évaluer dans quelle mesure ces données peuvent influencer l'étude des performances proches de zéro.

D. Conclusion de la partie 2 : Limites des mesures non paramétriques

L'étude de la littérature sur les seuils comptables menée dans la partie 1 de la thèse révèle que la méthodologie de mesure des irrégularités de distribution des résultats annuels publiés par les entreprises n'est pas définitivement arrêtée. Au contraire, un certain vagabondage méthodologique peut être observé, et plusieurs limites, mises en évidence notamment par Dechow, Richardson et Tuna (2003), McNichols (2003), Durtschi et Easton (2005) et Glaum, Lichtblau et Lindemann (2004) remettent en cause les interprétations des résultats publiés jusqu'alors.

La partie 2 de la thèse approfondit la recherche méthodologique en tentant de mesurer l'impact concret des principales limites identifiées : (A) Le choix de la méthode non paramétrique de mesure de l'irrégularité utilisée, (B) le choix de la largeur de l'intervalle d'observation, et (C) le choix du dénominateur utilisé pour mettre le résultat à l'échelle.

Chacun de ces trois axes d'étude a fait l'objet d'une présentation théorique (dans le chapitre IV). Puis leur impact concret a été étudié (dans le chapitre VI) à travers une étude empirique menée sur une population d'entreprises françaises cotées de 1992 à 2004. Les principaux résultats de ce travail sont les suivants :

- Le choix d'une méthode de mesure influence de manière importante les résultats des mesures, et par là même les conclusions d'une étude comparant les seuils. Il semble dès lors possible de revisiter les résultats obtenus jusqu'alors par les auteurs en comptabilité qui se sont intéressés au sujet. Par ailleurs, les méthodes par symétrie ou par moyenne arithmétique simple, pourtant les plus utilisées dans la littérature, apparaissent comme étant les plus fragiles. Elles rencontrent notamment des limites techniques qui interdisent de mesurer l'irrégularité pour certaines largeurs d'intervalle, et interdisent l'étude des années où sommet de la distribution et seuil sont proches l'un de l'autre. Les mesures par interpolation apparaissent donc plus appropriées mais soulèvent également des problèmes. Si le lissage permet d'éviter la volatilité des mesures par symétrie ou par moyenne des deux intervalles adjacents, une étendue de référence trop large risque de créer une difficulté pour représenter correctement l'endroit de la courbe où l'on se situe en

« aplatissant » toutes les données. Par ailleurs, les interpolations ne sont pas adaptées lorsque l'on se situe à proximité du sommet de la courbe. Enfin, le choix entre les différentes méthodes d'inter (ou d'extra) polation (linéaire, logarithmique, exponentielle) nécessite de faire une hypothèse sous jacente sur l'allure théorique de la distribution. En l'absence de postulat sur l'allure attendue, l'interpolation linéaire, qui fournit des mesures intermédiaires entre interpolation logarithmique et exponentielle, est la mesure la plus prudente.

- Le principal problème posé par l'utilisation d'indicateurs d'asymétrie est un problème conceptuel. Ces indicateurs considèrent en effet que l'effet de seuil se définit par la somme des deux effets : sur-représentation à droite, et sous-représentation à gauche. L'étude montre qu'il est utile de ne pas confondre ces deux phénomènes qui peuvent avoir des évolutions distinctes. Par ailleurs, l'indicateur d'asymétrie ne s'affranchit pas des problèmes de largeur d'intervalle. Ce problème est d'autant plus important que l'indicateur a été utilisé notamment pour comparer des populations de tailles différentes, et pour lesquelles la largeur d'intervalle retenue peut être différente. Les mesures sont faites sur la totalité des effectifs d'un intervalle, et non pas sur les effectifs irréguliers, ce qui peut diluer la mesure et provoquer des problèmes d'échelle. La mesure de l'asymétrie ne tient pas compte de l'endroit où l'on se trouve sur la courbe, et ne règle pas le problème de la pente non linéaire de la courbe de distribution, rendant toute interprétation délicate. Enfin, les indicateurs d'asymétrie peuvent être, par construction, totalement inopérant dans des cas extrêmes, notamment lorsque le dénominateur est nul.

- La largeur de l'intervalle a un très fort impact sur la mesure de l'irrégularité. Cela peut conduire jusqu'à inverser les résultats de certaines études. Cette limite est essentiellement liée à la démarche non paramétrique de mesure mise en œuvre dans la plupart des études réalisées jusqu'alors. Le fort impact du choix de la largeur sur la mesure a comme conséquence que le choix d'une largeur ad hoc n'est pas adapté à des études comparatives entre populations dont les caractéristiques sont différentes. L'utilisation d'une largeur théorique semble donc souhaitable même si l'étude montre qu'elle ne maximise pas nécessairement l'irrégularité. Cependant, l'influence de la largeur des intervalles sur les mesures conduit à

s'interroger sur l'intérêt de définir une largeur fixe des intervalles. Chercher à maximiser l'irrégularité semble plus pertinent dans l'objectif d'identifier la totalité des entreprises manipulatrices. La largeur de l'intervalle peut alors être considérée comme un indicateur mesurant une nouvelle dimension de l'irrégularité : le montant manipulé.

- Le choix du dénominateur utilisé pour mettre la variable résultat à l'échelle influence très sensiblement les mesures d'irrégularité. Il n'en demeure pas moins que, quel que soit le dénominateur utilisé, une irrégularité est presque toujours observable autour du seuil du résultat nul. Cette irrégularité est donc imputable à la variable résultat. Dans le cadre de la base de données, l'utilisation de l'actif total au dénominateur permet d'identifier les entreprises dont les performances sont proches de zéro avec les autres dénominateurs, alors que le contraire n'est pas toujours vrai. Pour cette raison, l'utilisation de cette variable est privilégiée. Enfin, la comparaison des différentes transformations de l'actif total montre que réduire par le total actif diminué du résultat évite certaines caractéristiques mathématiques indésirables liées à l'utilisation du total actif du début d'année.

En conclusion, il apparaît que si les études sont si sensibles à la largeur de l'intervalle d'étude et au choix de la méthode de mesure, cela est essentiellement dû à la nature non paramétrique des démarches mises en place jusqu'alors pour étudier les seuils comptables. Dès lors, proposer une méthodologie de mesure paramétrique des irrégularités constitue une contribution méthodologique substantielle. Elle permet en effet de s'affranchir des limites identifiées. Elle permet enfin de mesurer les montants manipulés par les entreprises qui évitent de publier des pertes. C'est ce qui est proposé dans la partie 3 de la thèse.

PARTIE 3 : PROPOSITION D'UN INSTRUMENT DE MESURE ET APPLICATION AUX ENTREPRISES FRANÇAISES COTÉES

L'étude de la littérature sur les seuils comptables menée dans la partie 1 de la thèse révèle que la méthodologie de mesure des irrégularités de distribution des résultats annuels publiés par les entreprises n'est pas homogène. Plusieurs démarches sont adoptées, et des limites méthodologiques sont évoquées. La partie 2 de la thèse montre que si les études sont si sensibles aux choix méthodologiques, cela est essentiellement dû à la nature non paramétrique des démarches mises en place jusqu'alors pour étudier les seuils comptables. La troisième partie de la thèse propose une réponse à ces limites.

La troisième partie est construite en trois temps. Le chapitre VII discute de l'opportunité d'adopter une démarche paramétrique de mesure des irrégularités. Une telle démarche suppose de connaître la loi de distribution théorique des résultats publiés (fonction mathématique). Cette loi étant jusqu'alors inconnue, une approche empirique est proposée pour l'ajuster.

Le chapitre VIII met en application cette démarche sur la base de populations d'entreprises françaises, allemandes et britanniques, cotées de 1992 à 2004. La fonction Pareto2 proposée pour ajuster les demi-distributions gauche permet de répondre aux difficultés étudiées dans la partie 2. Elle permet de sortir du dilemme du choix de la mesure (symétrie, interpolation ou extrapolation, linéaire, logarithmique ou exponentielle). Elle permet surtout de mesurer les irrégularités sur l'ensemble de la distribution, notamment en élargissant les intervalles d'étude, ce qui permet d'estimer le nombre total d'entreprises manipulatrices, et les montants manipulés.

Enfin, le chapitre IX met en œuvre la démarche paramétrique pour mesurer l'évolution des irrégularités de distribution des entreprises françaises de 1992 à 2004.

CHAPITRE VII - PROPOSITION D'UN NOUVEL INSTRUMENT DE MESURE DES IRREGULARITÉS

Après avoir distingué (section A) les méthodes paramétriques des méthodes non paramétriques dont les limites ont été examinées dans la partie 2, une discussion est menée sur la pertinence d'ajuster une loi de distribution des résultats (section B). L'hypothèse de Normalité des distributions est rejetée et une démarche empirique d'ajustement est proposée (section C).

A. Pourquoi faut-il adopter une méthode paramétrique ?

Parler d'irrégularités dans la distribution des résultats implique que l'on se réfère, consciemment ou non, à une distribution de référence considérée comme « régulière »[131]. En comparant la distribution empirique[132] à la distribution attendue, il apparaît d'éventuels écarts qui sont interprétés comme étant des irrégularités. Toutes les études sur les irrégularités dans les distributions de résultats font inévitablement référence à une distribution attendue des résultats.

Deux approches sont cependant possibles. La première n'est envisageable que si l'on connaît mathématiquement la loi de distribution. Il est alors possible de mettre en place des mesures dites « paramétriques » des irrégularités.

Dans la seconde approche, dite « non paramétrique », les paramètres de la loi de distribution dans son ensemble sont inconnus, mais on présuppose, quelle que soit cette loi, que la distribution est « lisse[133] ». Cette approche appelle plusieurs remarques :

- Elle est plus simple à mettre en œuvre dans la mesure où elle utilise peu l'outil mathématique.

[131] On parle aussi de « distribution attendue » ou de « distribution théorique ».

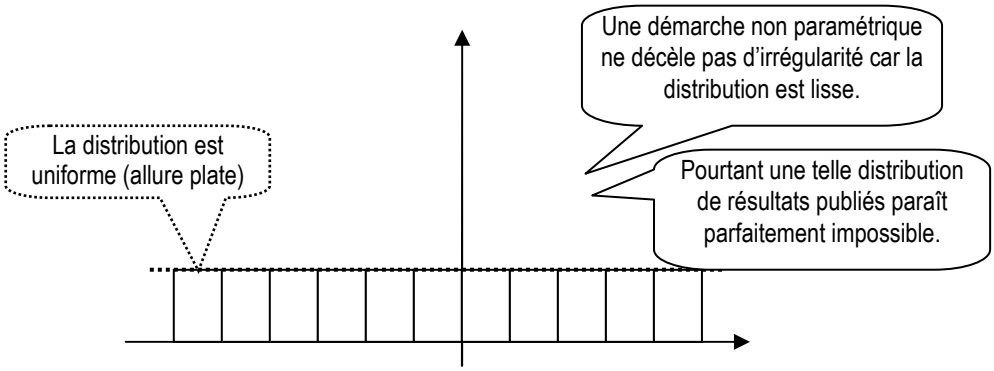
[132] Ou « réelle » ou « observée ».

[133] « Smooth » en anglais.

- Elles est plus robuste car les conditions sur lesquelles elle repose sont extrêmement peu contraignantes.

- Seules les discontinuités (c'est-à-dire les « pics » ou « ruptures ») de distribution sont prises en compte. Une allure « étrange » de la distribution ne sera pas considérée comme irrégulière tant que la distribution demeure « lisse ». Par exemple, une distribution uniforme des résultats (allure totalement plate, qui ne serait donc pas en forme de cloche) serait jugée régulière. La notion d'irrégularité est limitée aux discontinuités.

Figure 90 : Exemple de distribution lisse irrégulière



- Les irrégularités sont mesurées sans référence aux paramètres de la distribution. Elles sont mesurées à partir d'estimations locales, sur la base des effectifs observés de certaines classes isolées de l'ensemble de la population.

En définitive, les mesures non paramétriques ont les avantages de leurs inconvénients : simples et robustes, elles sont cependant imprécises, et ne reposent pas sur des fondements théoriques solides. La robustesse s'oppose ici à la fiabilité.

Dans la littérature en comptabilité, les chercheurs sont prudents. Ils ne cherchent pas à déterminer les paramètres de la loi de distribution des résultats. Ne connaissant pas cette loi, ils mesurent les irrégularités de manière non paramétrique. C'est le cas des méthodes examinées dans la partie 2 de la thèse. Les auteurs se contentent d'ajustements locaux en estimant une classe par les effectifs d'autres classes (extrapolations, interpolations, symétrie) ou en comparant les effectifs de deux classes successives (indicateurs d'asymétrie).

Cependant, si aucune étude ne se risque à proposer de manière explicite une loi de distribution des résultats, plusieurs études se fondent sur un second présupposé implicite : la distribution des résultats devrait avoir une forme gaussienne, c'est-à-dire

suivre une loi Normale. C'est le cas notamment de Burgstahler et Dichev (1997) qui mesurent les irrégularités par symétrie. Mesurer les irrégularités par symétrie implique une hypothèse sur un au moins des paramètres de la loi de distribution. Mard (2004) va plus loin puisqu'il ajuste les estimations en se référant à une allure normale de la loi de distribution. Dans les deux cas, les auteurs soulignent pourtant et paradoxalement le caractère non symétrique des distributions empiriques qu'ils observent.

En conclusion, les auteurs en comptabilité n'abordent pas le sujet de l'allure globale de la distribution des résultats de manière explicite, et lorsqu'ils le font, c'est de manière implicite et parfois contradictoire. L'objectif de cette troisième partie de la thèse est de mettre au point une mesure paramétrique des irrégularités, fondée sur une loi de distribution dont il faut découvrir les paramètres. Adopter une démarche paramétrique de mesure des irrégularités soulève deux types de problèmes :

1. Un problème technique de détermination d'une loi de distribution, qui sera l'objet du chapitre VIII.

2. Un problème théorique sur la nature des irrégularités qui est l'objet de ce présent chapitre. En effet, les discontinuités ne sont pas les seules manifestations d'irrégularités de distribution. D'autres irrégularités supposées ont déjà été identifiées, notamment les queues épaisses de distribution à gauche et les valeurs extrêmes négatives. Dans le cadre d'une recherche sur les irrégularités de distribution des résultats publiés, il ne peut donc pas être fait l'économie d'une réflexion plus générale sur l'ensemble des irrégularités potentielles qui pèsent sur ces distributions.

L'intérêt d'adopter une démarche paramétrique est d'effectuer des mesures beaucoup plus précises que toutes les mesures non paramétriques vues dans la partie 2. Une telle démarche s'affranchit des problèmes de largeur des intervalles étudiés dans la partie 2, et est indépendante de la taille des échantillons, ce qui permet plus facilement d'effectuer des comparaisons entre populations (études longitudinales ou internationales). Elle évite également le biais consistant à estimer un effectif irrégulier par l'utilisation d'un autre effectif irrégulier. Elle règle les problèmes de choix entre interpolation et extrapolation (linéaire, logarithmique, exponentielle). Elle permet des mesures indépendantes de l'emplacement où se situe l'irrégularité par rapport à l'ensemble de la distribution, ce qui règle notamment le problème des irrégularités proches du sommet. La limite de la méthode est bien entendu liée à ses qualités : elle repose sur la capacité d'ajuster par une loi mathématique la distribution des résultats de manière fiable.

B. Existe-t-il une loi de distribution des résultats ?

Avant de tenter d'ajuster une loi de distribution des résultats, chacune des irrégularités présumées des distributions des résultats est identifiée et discutée. Pour chacune, des explications économiques pouvant les justifier sont recherchées. Il apparaît alors que les discontinuités de distribution occupent un statut à part et qu'il est possible d'ajuster une loi de distribution des résultats indépendante de ces « pics ».

Pour identifier et discuter d'éventuelles irrégularités (autres que les discontinuités locales), il est nécessaire de se référer à une allure de référence. Ne connaissant pas cette distribution théorique, il est a priori impossible d'identifier ces irrégularités. Pour contourner ce problème, la réflexion est menée en deux temps. Dans un premier temps, la réflexion est menée sous l'hypothèse que la distribution suit une loi Normale[134]. La comparaison entre distributions réelles et loi Normale conduit à identifier des zones d'irrégularités potentielles qui sont discutées. Cette discussion conduit à considérer que seules les discontinuités situées autour du seuil du résultat nul sont des manifestations de gestion du résultat. Ce constat conduit alors à abandonner l'hypothèse de Normalité et à proposer un ajustement empirique de la loi de distribution (développé dans le chapitre VIII).

1. Les irrégularités sous hypothèse de Normalité

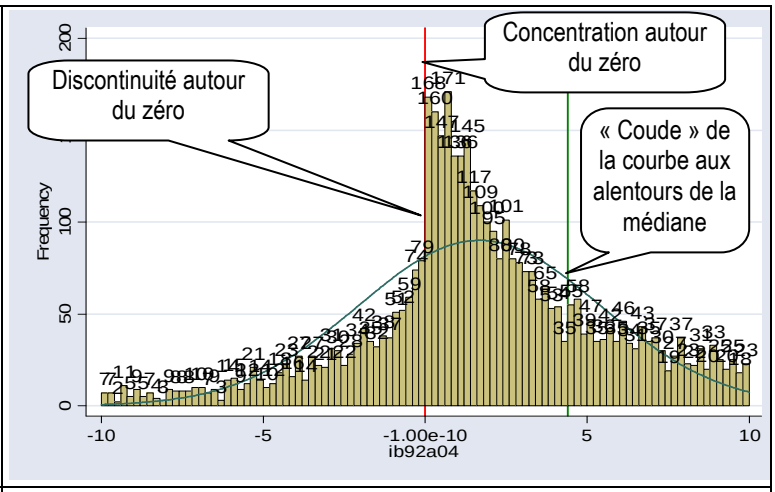
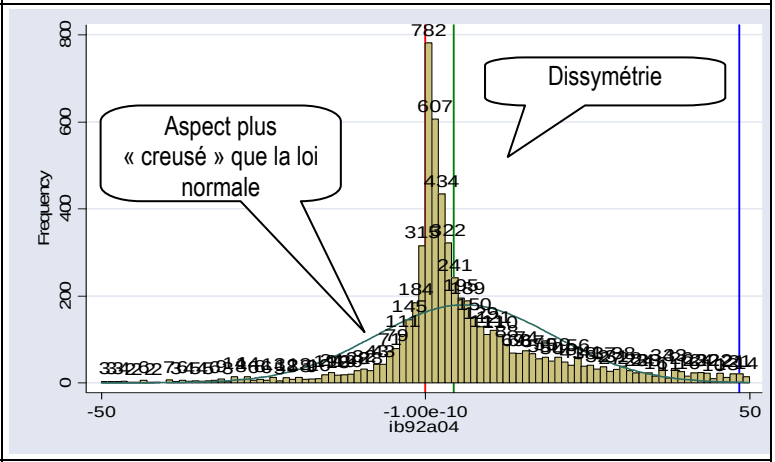
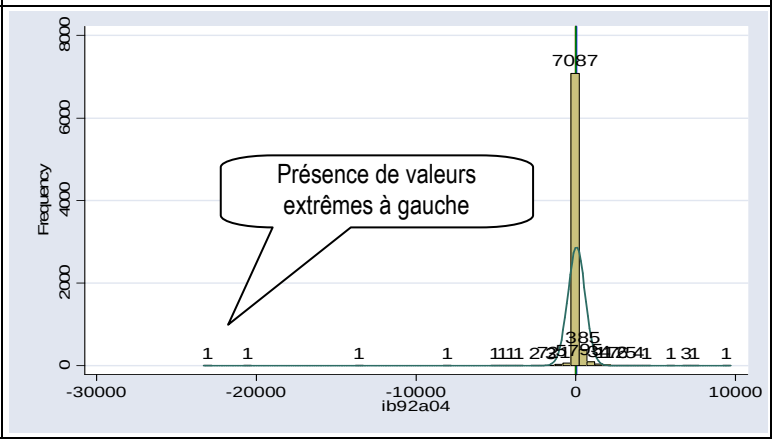
Si la distribution théorique suit une loi Normale, quels sont les écarts observés avec les distributions empiriques ? Ces écarts peuvent-ils être justifiés, ou faut-il les interpréter comme des irrégularités ?

Même si l'étude s'intéresse essentiellement à la variable « résultat mis à l'échelle par l'actif total », il n'est pas inintéressant de rappeler dans un premier temps l'allure des distributions de résultats non réduits. Il apparaît en effet dans les deux cas un certain nombre d'irrégularités communes.

[134] À la manière d'un raisonnement par l'absurde.

a. Distribution du résultat non réduit

Figure 91 : Distribution des résultats non réduits

Partie centrale de la distribution[135]	
Distribution sans les valeurs extrêmes[136]	
Distribution générale[137]	

[135] Sous Stata : histogram ib92a04 if ib92a04 > -10 & ib92a04 < 10, width (0.2) start (-10.0000000001) frequency normal blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)) xline(0,lcolor(red)) xline(48.43,lcolor(blue)) xline(4.405,lcolor(green)).

[136] Sous Stata : histogram ib92a04 if ib92a04 > -50 & ib92a04 < 50, width (1) start (-50.0000000001) frequency normal blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)) xline(0,lcolor(red)) xline(48.43,lcolor(blue)) xline(4.405,lcolor(green)).

La comparaison graphique (voir Figure 91) de la distribution de la variable résultat avec une loi Normale de même moyenne et écart type conduit à identifier les différences suivantes :

- La distribution ne présente pas une forme « en cloche » mais plutôt « en entonnoir renversé ». Une phase ascendante puis une phase descendante, mais pas de sommet « plat ».
- La distribution empirique est dissymétrique alors que la loi Normale est symétrique.
- La distribution empirique est plus concentrée (leptokurtique) que la loi Normale.
- La distribution empirique se caractérise par des queues plus épaisses de distribution (surtout à gauche[138]) que la loi Normale.
- Enfin, le mode de la distribution empirique n'est pas situé sur la moyenne, ni sur la médiane, mais sur le résultat nul.

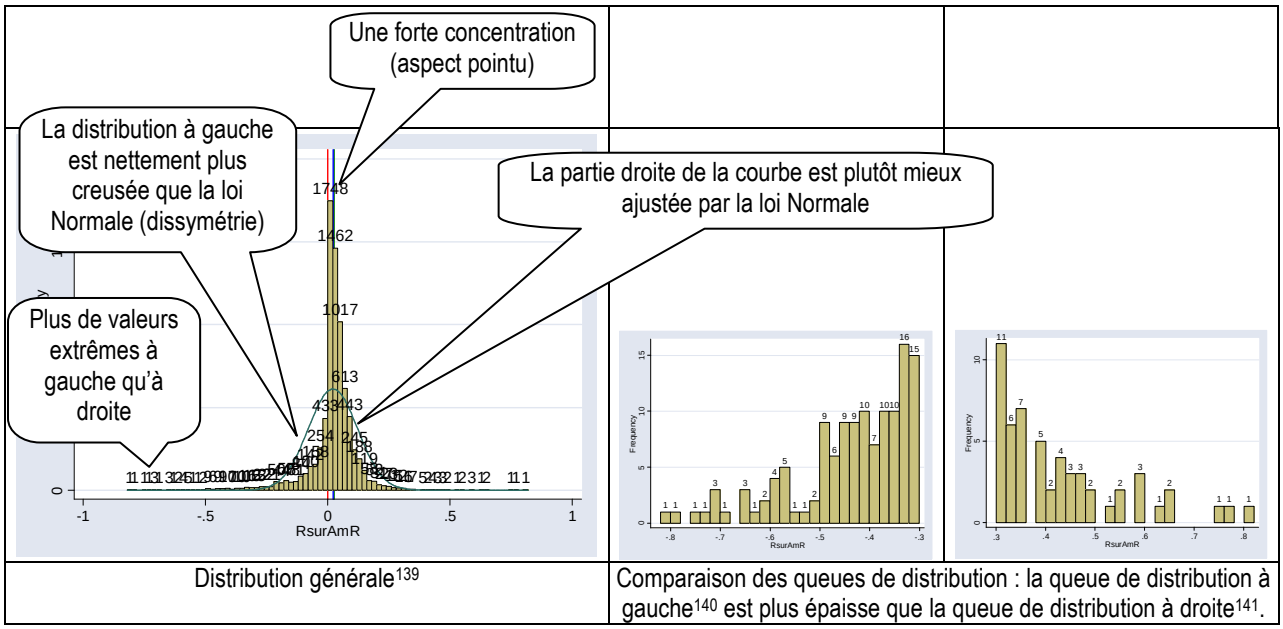
Mis à part la dernière, toutes ces différences sont constatées en observant également la distribution du résultat mis à l'échelle.

[137] Sous Stata : histogram ib92a04 , width (500) frequency normal blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)) xline(0,lcolor(red)) xline(48.43,lcolor(blue)) xline(4.405,lcolor(green)).

[138] En fait, la distribution est plus épaisse à droite, mais les valeurs extrêmes sont plus nombreuses à gauche. Donc lorsque l'on parle de queue épaisse de distribution à gauche, on parle des valeurs éloignées de la partie centrale de la distribution.

b. Distribution du résultat mis à l'échelle

Figure 92 : Distribution des résultats mis à l'échelle par la variable (AT-IB)



Distribution générale[139]

Comparaison des queues de distribution : la queue de distribution à gauche[140] est plus épaisse que la queue de distribution à droite[141].

La comparaison de la distribution empirique avec une loi Normale de même moyenne et écart type conduit à identifier les différences suivantes :

- La distribution ne présente pas une forme « en cloche » mais plutôt « en entonnoir renversé ».
- La distribution empirique est dissymétrique alors que la loi Normale est symétrique.
- La distribution empirique est plus concentrée (leptokurtique) que la loi Normale.
- La distribution empirique se caractérise par des queues plus épaisses de distribution à gauche que la loi Normale. La partie droite est mieux ajustée par la loi Normale.

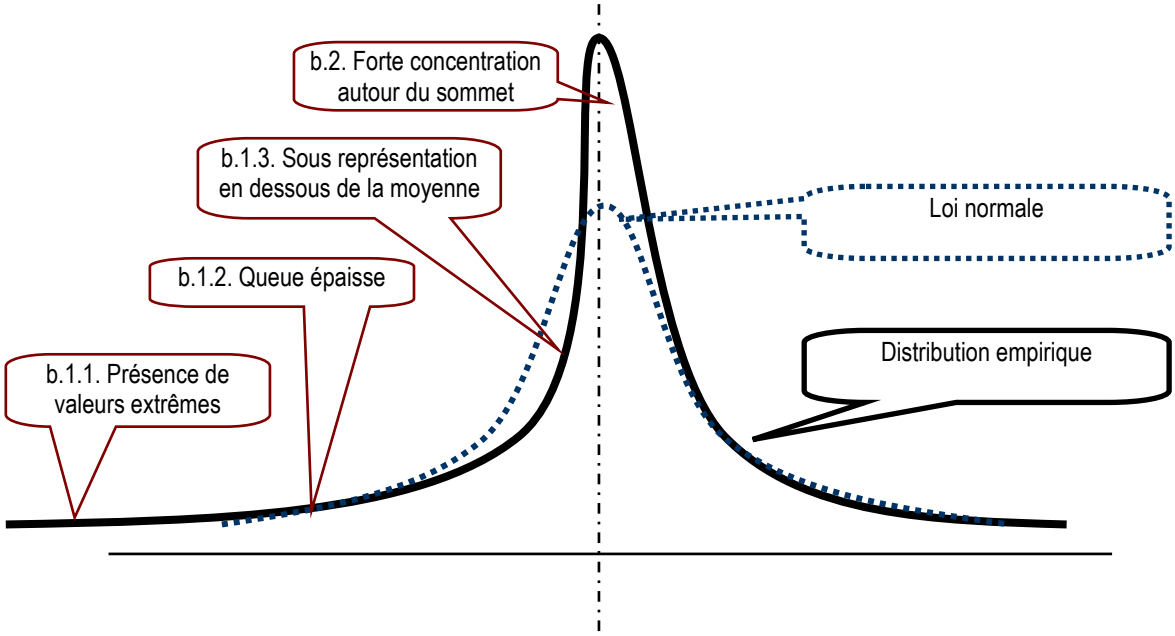
139 Fonction sous Stata : histogram RsurAmR , width(0.02) start (-0.82000000001) frequency normal blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)) xline(0,lcolor(red)) xline(0.0264,lcolor(green)) xline(0.0212111,lcolor(blue)).

140 Fonction sous Stata : histogram RsurAmR if RsurAmR<-0.3 , width(0.02) start (-0.82000000001) frequency blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)).

141 Sous Stata : histogram RsurAmR if RsurAmR>0.3 , width(0.02) start (0.3) frequency blcolor (black) blwidth (vvthin) addlabels addlabopts (mlabsize(medium)).

Ces différences sont représentées schématiquement sur la Figure 93. Elles sont ensuite successivement passées en revue.

Figure 93 : Représentation schématique des différences entre distribution empirique et distribution Normale



b.1. Dissymétrie

La distribution des résultats publiés est dissymétrique. Cette dissymétrie constatée en France est constatée également par Burgstahler et Dichev (1997) auprès des 76 000 observations de résultats de sociétés américaines. Plus précisément, trois sources de dissymétrie sont constatées : de fortes pertes, une épaisseur de la distribution à gauche, et des sous-effectifs à gauche de la moyenne.

b.1.1. Les fortes pertes

Il n'existe pas de limite économique à une perte, alors qu'il existe une limite à la hausse du résultat. En effet, le jeu de la concurrence dans le cadre d'une économie de marché implique qu'un avantage concurrentiel ne peut générer des rentes de situations de manière infinie. Les concurrents vont tenter de reproduire les ressources génératrices de rentes exceptionnelles. Par ailleurs, mécaniquement, générer un produit implique une activité consommatrice de ressources, donc générant elle-même des charges. On ne crée pas de la valeur à partir de rien.

Mais cette contrainte à la hausse n'existe pas à la baisse. Il est techniquement possible d'engager des charges de manière infinie sans générer de produits, ce qui conduit

mécaniquement à des pertes infinies. La contrainte du marché conduira cette entreprise à disparaître rapidement, mais durant quelques exercices (au moins un), l'entreprise peut afficher de très fortes pertes. De manière générale, une entreprise embourbée dans une spirale déficitaire peut réaliser des déficits extrêmement importants.

On ne peut donc pas considérer a priori la présence de fortes pertes comme une irrégularité de distribution liée uniquement à des pratiques de manipulation du résultat. C'est pour cette raison que l'étude des « grands bains comptables » est écartée dans le cadre de la thèse.

b.1.2. L'épaisseur de la distribution à gauche

Le principe de prudence peut être à l'origine de cette dissymétrie. Pour Givoly et Hayn (2002), la forte queue de distribution traduit la méfiance comptable (principe de prudence) vis-à-vis d'entreprises à risque, alors que le faible étalement à droite résulte de la résistance à traduire comptablement les bonnes performances. Mais la dissymétrie est-elle pour autant uniquement réductible à la prudence comptable ? Si la prudence s'applique à toutes les sociétés de la même manière, il faut s'attendre à observer une translation vers la gauche de la courbe de distribution des résultats, mais pas nécessairement de dissymétrie.

Cependant, les entreprises ne sont pas toutes nécessairement confrontées de manière semblable à des risques, et les entreprises dont les performances sont moins bonnes que les autres sont davantage confrontées à des événements défavorables. Dès lors, ces entreprises (dont le résultat est inférieur à la médiane) sont plus susceptibles de subir les conséquences du principe de prudence, et voient leur résultat davantage lesté de provisions. Cela peut expliquer que la moitié gauche de la courbe présente une pente plus rapide et s'étale davantage vers la gauche.

Le principe de prudence peut donc légitimement expliquer une dissymétrie, qui ne peut donc pas être réduite a priori à la manifestation de manipulations comptables.

b.1.3. Les sous-effectifs « juste à gauche » de la moyenne

Comme il vient d'être montré, le principe de prudence peut justifier la pente plus rapide à gauche de la moyenne. Par ailleurs, en considérant que les entreprises gèrent un portefeuille d'activités, elles sont susceptibles d'abandonner prioritairement les activités les moins rentables pour concentrer leurs efforts et leurs ressources sur les activités les plus rentables. Autrement dit, quand bien même la prudence comptable n'expliquerait pas

à elle seule la dissymétrie, un processus d'ajustement économique conduit les entreprises à quitter la partie gauche de la courbe de résultats en gérant leur portefeuille d'activités pour tendre à se rapprocher de la partie centrale de la courbe de distribution. La théorie de l'effort (Dechow, Richardson et Tuna, 2003) rejoint cette explication. Il n'est donc pas a priori anormal d'observer une dissymétrie.

b.2. Concentration

La distribution des résultats présente une forme « pointue », caractéristique d'une forte concentration. Cette concentration se fait autour de la médiane et/ou de la moyenne dans le cas de la variable mise à l'échelle, alors qu'elle se fait autour du résultat nul lorsqu'elle n'est pas mise à l'échelle. Ce constat interroge la nature économique du phénomène observé : le résultat.

b.2.1. Concentration autour du sommet[142]

Sous le postulat de normalité de la loi de distribution, la forte concentration doit être considérée comme une irrégularité. Autrement dit, les entreprises manipuleraient leurs résultats pour publier des résultats proches de la moyenne (médiane). La moyenne (médiane) annuelle des résultats publiés pourrait alors être interprétée comme un nouveau seuil, non encore évoqué dans la littérature.

On peut s'essayer à imaginer des justifications à cette hypothèse : afficher des résultats proches de la moyenne permet de donner une image de santé relative de l'entreprises, et il n'est pas impossible qu'il existe un coût à afficher des performances inférieures à celles de la plupart de ses concurrents. Les dirigeants pourraient alors être motivés pour atteindre ce seuil. Cette hypothèse permettrait également d'apporter une explication à la dissymétrie : les entreprises ayant des résultats supérieurs à la moyenne n'ont pas la même motivation à diminuer le résultat pour se rapprocher de la moyenne que les entreprises ayant des résultats inférieurs à la moyenne.

Cependant, le phénomène de concentration peut aussi s'expliquer par la nature non aléatoire de la distribution des résultats. (1) Il existe une certaine inertie des rentabilités. Une société fortement bénéficiaire une année peut difficilement être fortement déficitaire l'année suivante. Elle traversera des années progressivement de plus en plus difficiles avant d'atteindre une situation catastrophique. (2) Par ailleurs, dans un univers

[142] Nous étudions le sommet de la courbe (médiane et/ou moyenne) dans la section C du chapitre VIII. Il apparaît que si en général médiane et moyenne se confondent avec le sommet, certaines années les valeurs extrêmes peuvent avoir une forte influence sur le calcul de la moyenne.

économique concurrentiel, les lois du marché exercent une pression sur les entreprises tendant à homogénéiser leurs performances. (3) La théorie de l'effort déjà présentée dans la partie 1 peut également être invoquée pour justifier la concentration.

Il est donc naturel d'observer une concentration forte des distributions de résultats autour de la moyenne et ce phénomène ne peut être réduit à des manipulations comptables.

b.2.2. Concentration autour du zéro

Dans le seul cas du résultat non mis à l'échelle, une forte concentration est observée autour du zéro. Cela s'explique en grande partie par la composition de l'échantillon. Il y a beaucoup de petites entreprises et peu de grandes entreprises. Donc les petites entreprises réalisent des performances à hauteur de leurs investissements, que ce soient des pertes comme des bénéfices. Il peut donc sembler normal que les petites performances non mises à l'échelle soient concentrées autour du zéro.

Cependant, ce qui est plus difficilement explicable, c'est que (1) la pente à droite du zéro forme un « coude » aux alentours de la médiane. Autrement dit, rien ne justifie le surnombre d'entreprises entre le zéro et la médiane, ni le changement d'allure de la pente de la distribution entre ces deux valeurs par rapport au reste de la distribution. (2) Par ailleurs, il existe une rupture de la distribution entre résultat faiblement négatif et résultat nul. Cette discontinuité ne peut être justifiée par la composition de l'échantillon[143]. La petite taille devrait avoir le même effet du coté négatif et positif. Or la concentration n'est pas symétrique.

En résumé, s'il est difficile d'interpréter la distribution de résultat non réduit car l'effet de taille rend l'observation délicate, la forte concentration des seuls résultats positifs autour du zéro ne se justifie pas.

2. Théorème central limite

Les résultats des entreprises ne sont pas des données aléatoires, mais sont fondés sur l'activité réelle des entreprises et les décisions stratégiques plus ou moins pertinentes de leurs dirigeants. L'activité économique génère une rentabilité qui n'est pas aléatoire[144].

[143] Comme cela a déjà été étudié dans la partie 1, la théorie de l'effort ne peut justifier le « saut » au résultat nul puisque l'effort consenti par une entreprise en difficulté est réalisé avant de connaître le résultat de fin d'exercice. L'effort peut donc justifier une dissymétrie ou une concentration, mais en aucun cas une discontinuité autour d'un seuil précis.

[144] Quand bien même les résultats seraient aléatoires, cela ne signifierait pas pour autant qu'ils dussent suivre une loi Normale.

« Les entreprises ne jouent pas aux dés[145] », pourrait-on dire en parodiant Einstein. Chaque année, la distribution des résultats d'une population d'entreprises va donc présenter des caractéristiques dont il est impossible de présumer les caractéristiques a priori[146].

Le théorème central limite enseigne que si l'on agrège un grand nombre de tirages d'une variable suivant une même loi, la distribution agrégée tendra vers une loi Normale. Pour atteindre une telle normalisation de la distribution, il est nécessaire d'agréger un grand nombre de tirage, autrement dit, d'étudier la distribution des résultats des entreprises d'un pays sur plusieurs centaines d'années. Cela est actuellement impossible, et le sera toujours. Tout d'abord, de telles données ne sont pas disponibles. Mais surtout, la composition de la population étudiée et le contexte économique subissent des changements trop importants sur une longue période pour que l'on puisse accepter l'idée que la loi de distribution des résultats demeure inchangée sur le très long terme. Autrement dit, le théorème centrale limite ne peut être mis en application. La loi des distributions de résultats, agrégés sur plusieurs années ou non, est inconnue.

3. Synthèse et apports : La loi Normale n'est pas la loi de distribution

Tout d'abord, il est impossible d'appliquer le théorème central limite. Chaque année, la distribution des résultats présente des caractéristiques récurrentes (allure d'entonnoir renversé, dissymétrie, concentration). Or la dissymétrie et la concentration ne peuvent être réduits à des manifestations de manipulations comptables. Pour ces raisons, la distribution des résultats agrégés ne peut être assimilée à une distribution Normale.

La discontinuité au seuil du résultat nul occupe une place à part dans les irrégularités ; Elle ne peut se justifier économiquement. Pour mettre en œuvre une démarche paramétrique de mesure de ces discontinuités, il est donc nécessaire de trouver un meilleur ajusteur que la loi Normale. Il faut chercher une autre loi mathématique caractérisée par une allure asymétrique, concentrée, et continue (« lisse »).

[145] Indépendamment de l'omniprésence du facteur risque dans la plupart des décisions.

[146] Une façon de tester empiriquement l'allure des résultats non manipulés serait de réaliser des expérimentations en laboratoire, par exemple à partir d'un jeu d'entreprise. Les participants ne prendraient aucune décision comptable. L'information comptable ne pourrait dès lors pas être manipulée. Quelle serait l'allure de la courbe de distribution des résultats ?

C. À la recherche d'une loi de distribution

Devant l'incapacité de prédire a priori la nature de la loi de distribution, l'étude adopte une démarche empirique originale : observer les distributions des résultats publiés, pour en déduire une allure caractéristique (loi de distribution théorique) des résultats annuels des entreprises françaises. La démarche de recherche se fonde sur les propositions suivantes :

1. Il est possible d'ajuster empiriquement une loi de distribution des résultats publiés.
2. Cette loi (l'allure de la distribution) est invariable (aux paramètres près) dans la population de référence (entreprises françaises[147]), quelle que soit l'année, et la taille de l'effectif. C'est une loi économique.

Cette section précise les fondements de cette démarche empirique, et les modalités de sa mise en oeuvre.

1. Fondements de la démarche d'ajustement empirique de la loi

Les différences observées en comparant l'allure des distributions empiriques à une distribution normale ont été discutées. L'existence de la dissymétrie, la sur-représentation des fortes pertes, la queue épaisse à gauche, la forte concentration autour de la moyenne peuvent être justifiées économiquement. Seules les discontinuités autour du seuil du résultat nul peuvent être considérées comme des manifestations de pratiques de manipulations comptables.

Si la loi Normale ne peut être retenue pour estimer la loi théorique de distribution, il est nécessaire de découvrir une loi plus adaptée aux caractéristiques de la population. Découvrir une loi de distribution des résultats a d'importantes conséquences en termes d'interprétation. En effet, une fois que la loi est déterminée, tout écart par rapport à cette loi devient une irrégularité, et est donc interprétée comme la manifestation de manipulations comptables. La difficulté, c'est qu'il n'existe pas a priori de théorie permettant de définir la loi de distribution des résultats.

Dans un papier de recherche non encore publié, Chen, Lin, Wang et Wu (2005) mettent en œuvre une méthodologie paramétrique de mesure des irrégularités de résultat qui

[147] Nous montrons dans la section 3-II-D que l'allure est la même pour les entreprises britanniques et allemandes.

permet de mesurer la « largeur[148] » (interprétée comme les montants manipulés) et la « hauteur[149] » (interprétée comme la motivation à manipuler). Les auteurs estiment que les entreprises sont de deux types, et en déduisent qu'il faut utiliser un mélange de deux loi normales. Mais ce postulat n'est pas argumenté[150]. Or de ce postulat découlent les mesures des irrégularités, et les résultats de l'étude sont totalement dépendants de ce postulat. Il y a là un raisonnement en quelque sorte tautologique. Une simple observation visuelle des distribution conduit à constater que les distributions de variation de résultat, ou d'erreurs de prévisions sont plus concentrées, et de telles distributions seront donc automatiquement plus éloignées de la distribution de référence normale mixte. Cette étude confirme, et la difficulté de déterminer une loi théorique de distribution des résultats, et l'intérêt d'une démarche empirique.

L'hypothèse de Normalité de la distribution étant rejetée, comment faire pour déterminer la loi de distribution des résultats non manipulés alors que l'observateur ne dispose que des distributions empiriques, donc potentiellement manipulés ?

L'observation graphique (visuelle) des distributions met en évidence des similitudes dans leur allure générale, quels que soient l'année ou le pays retenu. Cette courbe semble toujours plus pointue qu'une loi normale, et est dissymétrique. Intuitivement, il semble intéressant de tenter d'exploiter ces distributions de résultats publiés (donc potentiellement manipulés) afin d'en déduire une approximation de la loi de distribution théorique des résultats (non manipulés). L'intuition se fonde sur l'idée (a) qu'un nombre suffisamment grand d'observations d'une part, associé (b) à un traitement mathématique adéquat d'autre part, permet de surmonter ce problème.

a. L'utilisation d'une caractéristique comptable

L'originalité de la démarche proposée dans cette étude consiste à utiliser une propriété comptable : sur le long terme, les manipulations comptables tendent mécaniquement à s'annuler. Une manipulation à la hausse une année doit inévitablement être compensée par une manipulation à la baisse dans les années ultérieures. Autrement dit, si les

[148] Les auteurs parlent de « magnitude ».

[149] Les auteurs parlent de « frequency ».

[150] Les auteurs écrivent : « Nous postulons qu'il y a deux types d'entreprises. Le résultat est une variable aléatoire dont la moyenne est $\mu 1$ et la variance $\sigma 1^2$ pour le premier type et dont la moyenne est $\mu 2$ et la variance $\sigma 2^2$ pour le second type. La probabilité qu'une firme soit de type un ou deux sera ρ ou $1-\rho$ ». Notre traduction de : « We assume that there are two types of firms in the economy. Earning is a normal random variable with mean $\mu 1$ and variance $\sigma 1^2$ for the first type and with mean $\mu 2$ and variance $\sigma 2^2$ for the second type. The probability for a firm to be type one or two will be ρ or $1-\rho$ ».

distributions empiriques sont manipulées, sur une longue période, l'agrégation des distributions doit conduire à compenser ces manipulations et estomper les irrégularités de distribution. Autrement dit, la longueur de la période d'étude, sans être suffisamment longue pour permettre l'utilisation d'une loi Normale, peut justifier la mise en œuvre d'une procédure d'ajustement empirique de la distribution agrégée.

b. Le traitement mathématique adéquat

Mathématiquement, la démarche qui consiste à découvrir une loi de distribution théorique à partir d'une distribution empirique s'appelle un « ajustement ». De nombreux logiciels permettent d'effectuer de tels ajustements. Ils possèdent chacun un catalogue de lois. Ils mesurent pour chaque loi un indicateur de qualité de l'ajustement. Un classement des lois est effectué, et il est possible d'en déduire la loi théorique (parmi les lois présentes dans le catalogue) permettant le meilleur ajustement de la distribution empirique.

Mais l'ajustement ne peut pas être réalisé à partir de la distribution agrégée des résultats (mis à l'échelle) bruts. En effet, la loi recherchée est la loi de distribution des résultats annuels non manipulés. Or chaque année, les résultats des entreprises correspondent à un « tirage » indépendant. Par ailleurs, il est important de souligner que la loi recherchée est la loi de distribution théorique des résultats attendus, c'est-à-dire la loi de distribution des résultats réels (non manipulés) des entreprises. La loi théorique peut donc être différente de la distribution empirique puisque les résultats publiés sont considérés comme potentiellement manipulés.

Le traitement mathématique doit donc au préalable résoudre ces deux problèmes : (1) il doit être conduit toutes choses égales par ailleurs, donc en annulant au préalable l'influence de l'environnement sur les performances économiques, et (2) pour parvenir à ajuster une loi théorique à partir d'une loi empirique potentiellement irrégulière, il faut identifier les sources d'irrégularités potentielles, puis traiter ces « zones inquiétantes ».

La solution proposée est de réduire et centrer les distributions avant de les agréger. La sous-partie suivante (mise en oeuvre de la démarche) précise pourquoi cela répond aux deux problèmes soulevés.

2. Mise en œuvre de la démarche

Pour mettre en œuvre la démarche, deux problèmes doivent être étudiés préalablement : (1) l'exploitation de données pluriannuelles[151] pour déterminer une loi de distribution annuelle, et (2) le traitement des données supposées irrégulières.

a. Utilisation de données pluriannuelles

La base de données est constituée des résultats mis à l'échelle des entreprises françaises cotées de 1992 à 2004. Or l'étude cherche à ajuster la loi de distribution des résultats annuels. Comment estimer une loi annuelle à partir de données pluriannuelles ? Faut-il ne retenir qu'une année ? Si oui, laquelle ? Pour ajuster une loi, il est préférable de maximiser le nombre d'observations. Il est donc souhaitable d'ajuster la loi de distribution des résultats à partir de l'ensemble de l'échantillon : 1 009 sociétés sur 13 ans (soit 8 072 données en tenant compte des entrées et des sorties de cote).

Chaque année, l'environnement et la conjoncture économique peuvent influencer les performances des entreprises. Les paramètres de la distribution annuelle (moyenne, écart type) peuvent donc être différents d'une année sur l'autre (la performance moyenne peut être moins élevée en période de crise, les résultats plus étalés selon le contexte…). Mais déterminer une loi de distribution des résultats repose sur le postulat que cette loi (l'allure de la courbe) est toujours la même. Elle est considérée comme étant une donnée économique, identique quelle que soit l'année ou le pays étudié, aux paramètres près (moyenne et écart type varient en fonction de la conjoncture). La loi recherchée est donc la loi de distribution annuelle, et non pas la loi des distributions agrégées.

Les résultats étant publiés annuellement, les 13 distributions de résultats annuels sont assimilables à 13 « tirages » indépendants, dont les caractéristiques (moyenne/médiane et écart type) sont différentes. Il n'est donc pas souhaitable d'additionner ces distributions telles quelles, comme l'ont fait les précédents chercheurs, pour ajuster une loi théorique de l'allure générale des distributions. Pour « homogénéiser » les données, il faut centrer et réduire chaque distribution annuelle. Cela permet d'obtenir 13 distributions centrées sur la moyenne/médiane annuelle, et de même écart type 1. Ce sont ces 13 distributions transformées qui peuvent être agrégées. Et cette distribution agrégée permet alors de

[151] La variable étudiée est le résultat (IB dans la base Global) réduit par l'actif diminué du résultat (AT-IB). Les qualités de cette variable ont été développées dans la partie 2.

tracer une allure de la distribution indépendamment des paramètres annuels qui pourra ensuite être ajustée par une loi mathématique.

L'étude de l'agrégation des distributions centrées-réduites a donc deux avantages : (1) en permettant d'agréger les données de plusieurs années, elle maximise le nombre d'observations et augmente la qualité des traitements statistiques, et (2) en lissant les irrégularités autour du résultat nul, elle permet un traitement de la zone inquiétante. Ce deuxième point est développé dans la sous-partie suivante.

b. Traitement des données supposées irrégulières (« zone inquiétante »)

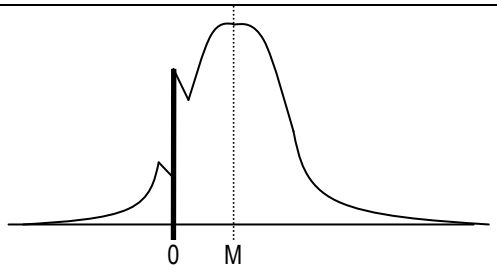
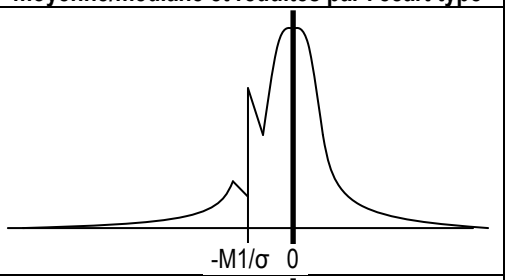
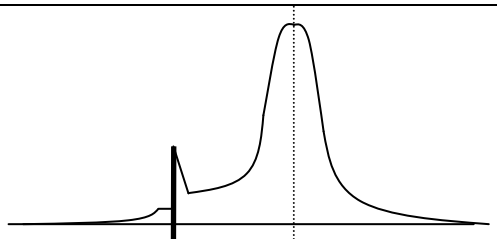
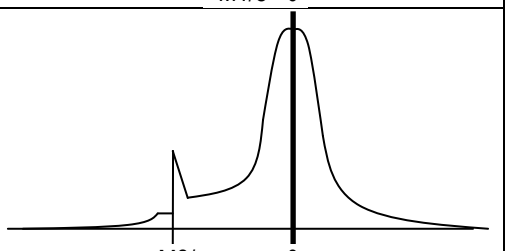
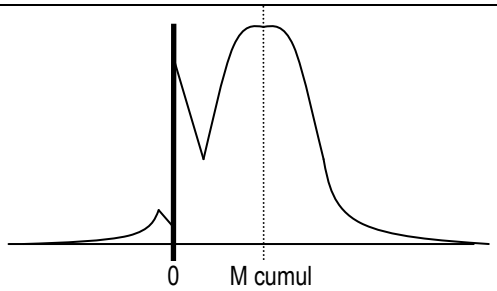
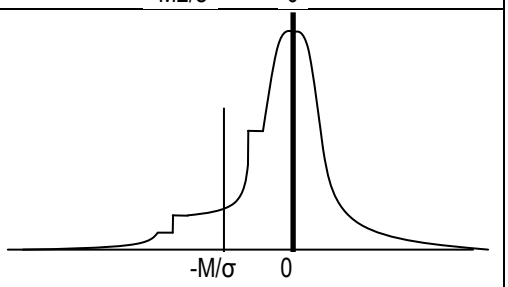
Comment ajuster une loi théorique régulière à partir de données dont on suppose qu'une partie est irrégulière ? Autrement dit, comment neutraliser la zone inquiétante ?

Pour ajuster une loi mathématique à partir d'une distribution empirique dont une zone est présumée irrégulière, il faut en théorie supprimer cette zone et la remplacer par exemple par des valeurs extrapolées à partir du reste de la distribution. Une telle démarche est délicate car elle suppose de connaître les données irrégulières, et une extrapolation conduirait à utiliser des estimations non paramétriques que l'étude cherche justement à dépasser.

La solution proposée est de jouer sur l'effet du grand nombre d'observations. En centrant et en réduisant les données, le point invariant n'est plus le résultat nul. La moyenne (comme la médiane) des distributions varie chaque année. Donc les irrégularités ne sont plus cumulées. Sur un grand nombre d'années, les irrégularités s'estompent. Le graphe ci-dessous (Figure 94) illustre cette propriété. La première colonne représente de manière schématique l'agrégation de deux distributions brutes[152]. Les moyennes M1 et M2 sont différentes. La distribution agrégée est donc centrée sur une moyenne $M = (M_1+M_2)/2$. Les irrégularités autour du résultat nul se cumulent car le point zéro est le point invariant des distributions. La deuxième colonne représente schématiquement l'agrégation de distributions centrées réduites. Les deux distributions sont donc centrées sur la moyenne = zéro, et les irrégularités sont respectivement observables à deux endroits différents ($-M_1/\sigma_1$ et $-M_2/\sigma_2$). La distribution agrégée ne cumule donc pas les irrégularités sur le même point, mais les décale l'une par rapport à l'autre. Elles tendent donc à s'estomper.

[152] Les données dites « brutes » correspondent aux résultats mis à l'échelle par un indicateur de taille (Total actif) mais non centrés-réduits. Il ne faut pas confondre la « réduction » (ou mise à l'échelle) par un indicateur de taille, de la réduction par l'écart type. Ces deux réductions sont faites successivement.

Figure 94 : Transformation de la zone irrégulière lors de l'agrégation des données ; comparaison entre agrégation des données brutes et centrées-réduites

	Agrégation des distributions sans transformation	Agrégation des distributions centrées sur la moyenne/médiane et réduites par l'écart type
Année 1		
Année 2		
Agrégation		
	Le résultat nul est le point invariant. Les irrégularités se renforcent lorsque l'on agrège les distribution.	Les courbes sont centrées et réduites. Les irrégularités au résultat nul ne se cumulent plus et ont tendance à se neutraliser.

Ce phénomène conduit donc à penser que la suppression de la zone inquiétante n'est pas nécessaire pour ajuster la loi de distribution. L'ajustement de la distribution empirique (des résultats potentiellement manipulés) sur plusieurs années, après centrage et réduction permet d'établir une approximation raisonnablement fiable de la loi de distribution théorique des résultats (non manipulés). In fine, il est inutile de retirer les zones inquiétantes et l'ajustement est effectué directement à partir de la somme des distributions centrées réduites.

3. Caractéristiques de la loi attendue

a. Caractéristiques attendues

La loi de distribution attendue doit réunir deux caractéristiques.

a.1. Une allure « en entonnoir renversé »

La forme en « entonnoir renversé » est caractéristique d'une forte concentration de la distribution autour du sommet. La concentration de la distribution des résultats (mis à l'échelle ou non) est beaucoup plus forte que celle obtenue avec une loi normale. Cette concentration ne s'explique pas par la présence de variables extrêmes. Il est donc indispensable de tester des lois d'ajustement dont l'allure est plus « pointue » que la loi normale.

a.2. Une dissymétrie

L'observation visuelle des distributions, et la mesure des Skewness montrent que la distribution des résultats annuels est dissymétrique, très étalée, mais aussi plus creusée sur la partie gauche (ascendante). Il est donc indispensable de tester des lois d'ajustements dissymétriques.

a.3. Une courbe non bornée

La distribution des résultats mis à l'échelle par (actif – résultat) est bornée sur l'intervalle $]-1 ; +\infty[$. Mais la distribution des résultats mis à l'échelle puis centrés et réduits n'est pas bornée. Elle est définie sur $]-\infty ; +\infty[$. La loi recherchée doit donc être définie sur $]-\infty ; +\infty[$.

b. Lois testées

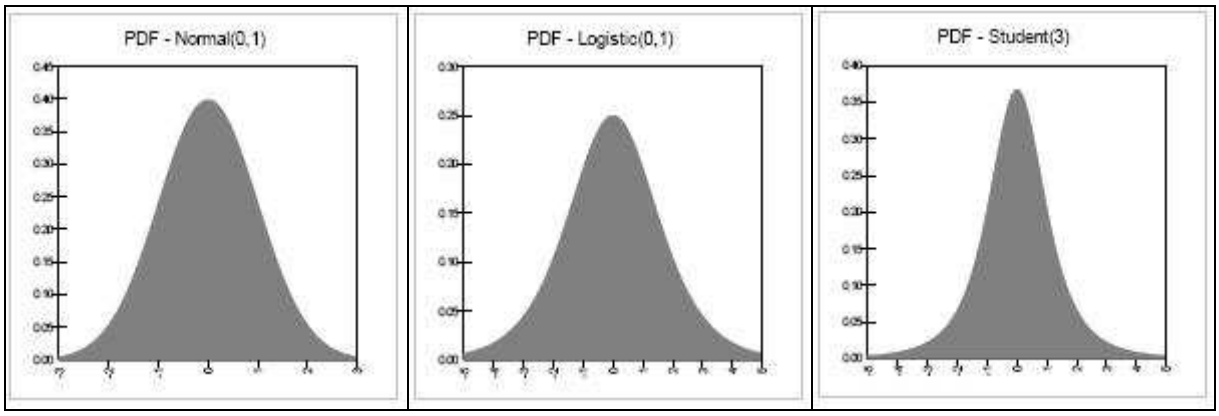
Deux séries de lois sont successivement testées. La première série sert à ajuster l'ensemble de la distribution. La seconde sert à ajuster les demi-distributions.

b.1. Ajustement global

Les lois en forme « de cloche » et/ou « d'entonnoir renversé » peuvent être réunies en deux familles :

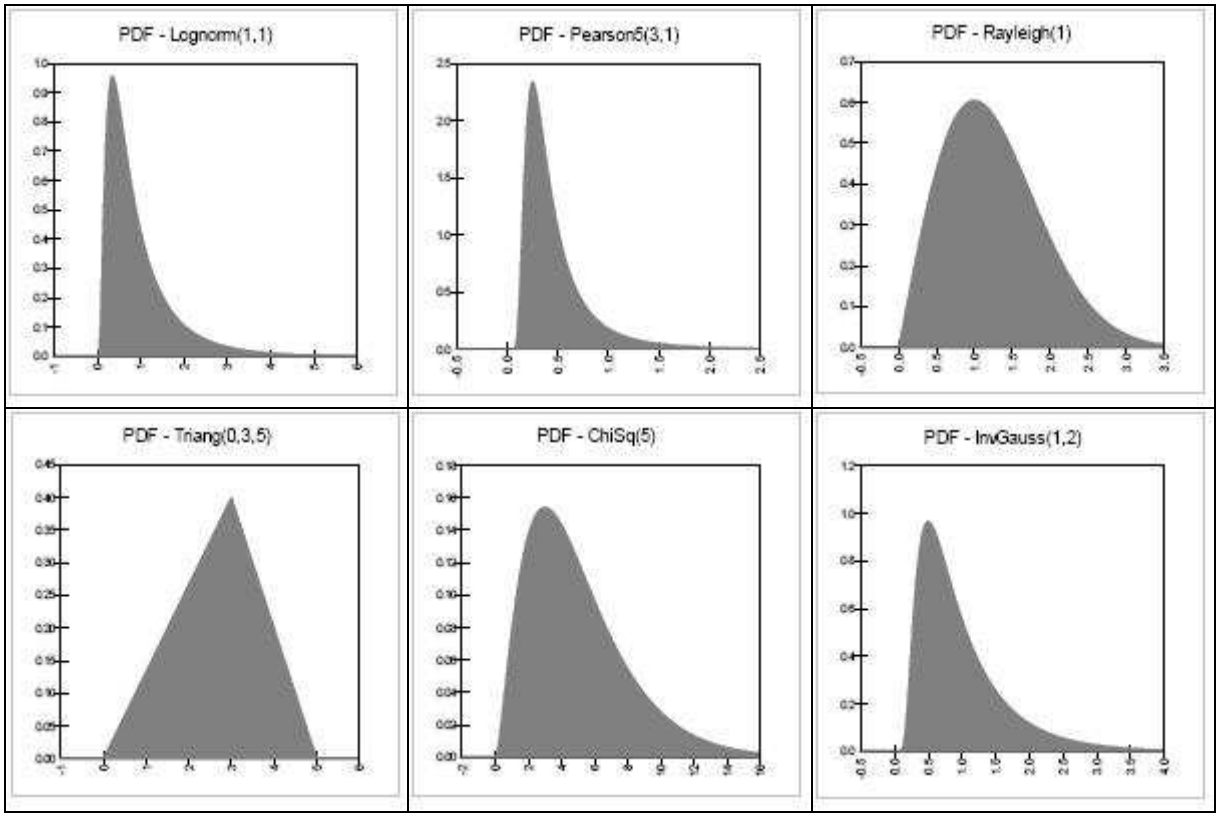
1. Les lois symétriques (Laplace, Logistique, Normale, Student). Par définition, ces lois ne permettent pas de remplir la condition d'asymétrie.

Figure 95 : Exemples de lois de distribution symétriques



2. Les lois dissymétriques (Béta, Chi2, Erlang, Gamma, Inverse Gaussienne, LogLogistique, Lognormale, Pearson, Rayleigh, Triangulaire, Weibull). Répondant à la condition d'asymétrie, ces lois sont cependant bornées à gauche. Elles ne sont pas définies sur]- ∞ ; +∞[. Ces lois ne permettent donc pas de remplir la troisième condition requise.

Figure 96 : Exemples de lois de distribution dissymétriques

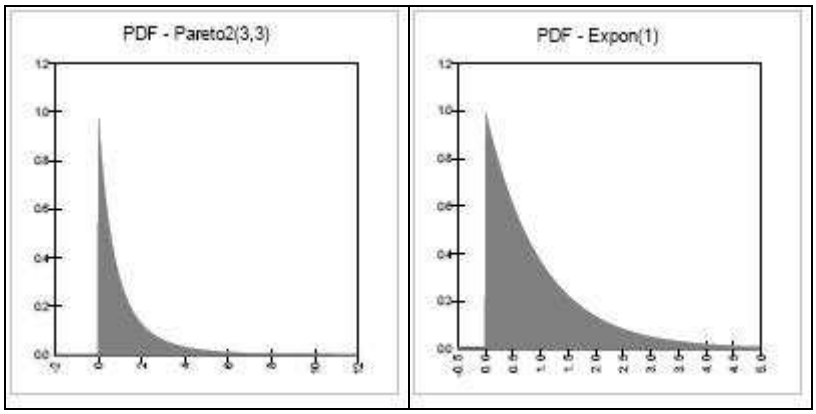


Il est donc difficile de trouver une loi d'ajustement qui remplisse toutes les conditions. Pour faire face à cette difficulté, l'étude a procédé en deux temps. Dans un premier temps, la distribution a été ajustée au mieux avec les lois disponibles citées ci dessus. Cette première étape est appelée « ajustement global » (c'est-à-dire qu'il est réalisé sur la totalité de la distribution). Dans un deuxième temps, il a été procédé à un ajustement partiel.

b.2. Ajustement partiel

Comme l'ajustement global utilise le catalogue des lois dont les limites ont été soulignées (dans la sous-partie b.1), une deuxième étape permet d'améliorer les prévisions, et de prendre en compte la dissymétrie des distributions. La distribution est découpée en deux et chacune de ces demi-distributions est ajustée indépendamment. La partie ascendante d'une part (à gauche du sommet) et la partie descendante d'autre part (à droite du sommet) sont ajustées successivement par une loi monotone (Exponentielle, Pareto), c'est-à-dire une loi ne connaissant pas d'inversion de pente (donc pas de forme en « cloche » ni en « entonnoir renversé »).

Figure 97 : Exemples de lois monotones



Il faut alors distinguer deux équations différentes, une pour chaque demi-distribution (à gauche et à droite du sommet). Les deux lois doivent normalement se rejoindre en leur milieu. Cependant, l'étude de la base de données a permis d'observer que les années pour lesquelles la performance moyenne des entreprises est négative sont rarissimes[153], et la médiane n'est jamais négative. L'irrégularité du zéro se produit donc chaque année sur la moitié gauche de la courbe. Il est donc envisageable de n'ajuster que cette moitié

[153] Une fois en 13 ans dans notre population d'entreprises françaises.

gauche de la courbe. Cela est suffisant pour effectuer les estimations d'effectifs attendus sur les intervalles étudiés (au voisinage de zéro), et donc de mesurer les irrégularités.

c. Outils statistiques

c.1. Logiciel utilisé

Il existe plusieurs logiciels d'ajustement de loi. Le plus complet est, à notre connaissance, le logiciel « Best-Fit » édité par la société « Palissade ». Ce logiciel fait partie de la suite « @risk », logiciel d'analyse du risque[154]. Le catalogue des lois testées par le logiciel est le suivant :

Tableau 63 : Catalogue des lois testées par le logiciel Best-Fit

Béta	Error Function	Inverse gaussienne	Pearson Type V
Béta (généralisée)	Erlang	Logistique	Pearson Type VI
Béta (subjective)	Exponentielle	Log-logistique	Pert (Béta)
Binomiale	Extreme value	Log-Normale format 1	Poisson
Chi²	Gamma	Log-Normale format 2	Rayleigh
Cumulative (ascendante)	General	Negative binomiale	Student's t
Cumulative (descendante)	Géométrique	Normale	Triangulaire
Discrète	Hypergéométrique	Pareto (first kind)	Uniforme
Discrète uniforme	Integer Uniform	Pareto (second kind)	Weibull

c.2. Critères de qualité des ajustements

Trois tests sont utilisés pour mesurer la qualité des ajustements.

La statistique du Chi²

Le test du Chi² est le plus connu des tests d'ajustement. Il est normalement utilisé dans le cas de données discrètes, mais peut être utilisé également dans le cas de données continues. C'est le cas dans cette étude qui mesure la qualité de l'ajustement d'une fonction mathématique continue à une distribution empirique. Pour mesurer la qualité d'ajustement d'une loi continue, il est nécessaire de découper l'axe des abscisses en intervalles, transformant la fonction continue en succession de classes[155], donc en données discrètes.

[154] Depuis 2008, le logiciel Best-Fit est totalement intégré dans le logiciel @risk et n'est plus disponible séparément.

[155] En anglais (et dans le logiciel) : « bins ».

La faiblesse du test du Chi² est qu'il n'y a pas de règle pour sélectionner le nombre et le découpage des classes. Dans certaines situations, il est possible de formuler des conclusions différentes à partir des mêmes données, en fonction de la manière dont sont définis les intervalles. L'arbitraire dans la sélection des intervalles peut être supprimé en demandant au logiciel d'utiliser des intervalles équiprobables. Le logiciel ajuste la taille des intervalles sur la base de la distribution ajustée, de telle sorte que chaque classe contienne la même probabilité d'effectifs.

L'adéquation de la distribution empirique à la distribution théorique est mesurée par la « distance du Chi-deux » qui est une moyenne pondérée d'écarts quadratiques entre les deux distributions.

La statistique de Kolmogorov-Smirnov

Le second test d'ajustement utilisé dans le cadre de fonctions continues est le test de Kolmogorov-Smirnov (K-S). L'adéquation de la distribution empirique à la distribution théorique est mesurée par la distance de Kolmogorov-Smirnov. À la différence de la distance du Chi-deux, cette distance n'est pas une moyenne pondérée des écarts, mais la valeur de la distance au point où les deux distributions sont les plus éloignées.

L'intérêt de ce test est qu'il ne nécessite pas de diviser la distribution en intervalles. Cela le rend moins arbitraire que le test du Chi-deux. Cependant, la faiblesse de ce test est d'accorder une plus grande importance à la partie centrale de la distribution. Il ne détecte en général pas les écarts d'ajustement qui interviennent sur les queues de distribution.

La statistique de Anderson-Darling

Le troisième test utilisé dans le cadre d'une fonction continue est le test d'Anderson-Darling (A-D). Comme le test de K-S, il ne nécessite pas de découpage de la fonction en intervalles. Mais à l'inverse du test K-S qui se focalise sur le centre de la distribution, le test A-D met l'accent sur les différences entre les queues de la distribution empirique et celles de la distribution ajustée.

Significativité[156] et valeur critique

Dans les trois cas, plus la mesure de la distance est petite, plus l'ajustement est bon. La significativité (P) mesurée par le logiciel est la probabilité qu'un échantillon tiré à partir de la loi théorique conduise à une distance plus importante que celle mesurée avec

[156] Dans le logiciel : « P-Value ».

l'échantillon empirique. Autrement dit, plus P est petit, moins l'ajustement est bon. Plus P est proche de 1, plus l'ajustement est bon.

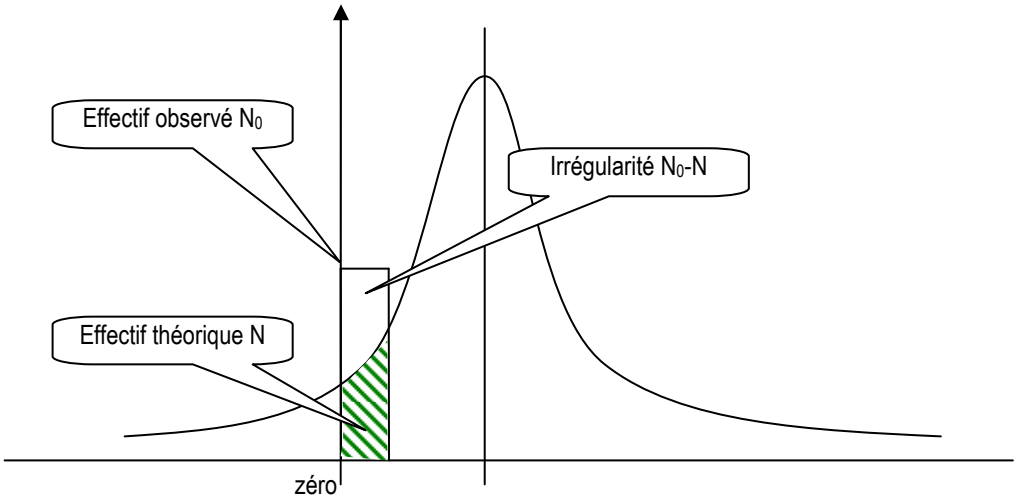
La valeur critique est la distance maximale acceptable pour que la probabilité (1-P) de rejeter une loi d'ajustement à tort soit inférieure à un degré de significativité fixé arbitrairement (par exemple 5 %). Dans l'étude, c'est la loi qui ajuste au mieux la distribution qui est recherchée. Aucune valeur critique n'est donc fixée, mais les lois sont classées en fonction de leur significativité P.

4. Finalité de la démarche

Dès lors que la distribution théorique est connue, mesurer l'irrégularité consiste très simplement, pour n'importe quelle largeur d'intervalle, et à n'importe quel endroit de la distribution, à faire la différence entre effectif théorique et effectif observé. L'effectif théorique est égal à la surface définie par la courbe et l'axe des abscisses entre les deux limites de l'intervalle considéré (Calcul d'intégrale réalisé directement par le logiciel).

L'intérêt de cette proposition est d'aboutir à une mesure paramétrique des irrégularités. Il devient alors possible de s'affranchir des limites méthodologiques constatées dans la partie 2 de la thèse (dépendance à la largeur des intervalles, myopie des indicateurs d'asymétrie, volatilité des mesures non paramétriques).

Figure 98 : Représentation graphique de la mesure des irrégularités par une méthode paramétrique



Par ailleurs, une telle méthode permet une représentation visuelle des irrégularités, de leur amplitude et de leur persistance, comme le montre le chapitre suivant.

CHAPITRE VIII - MISE EN ŒUVRE : AJUSTEMENT D'UNE LOI DE DISTRIBUTION DES RÉSULTATS DES ENTREPRISES FRANÇAISES

Pour estimer une loi de distribution théorique des résultats annuels des entreprises, l'étude procède de la manière suivante : (A) les valeurs extrêmes sont supprimées, (B) les distributions annuelles sont réduites, (C) centrées, et (D) la distribution est ajustée.

Il est indispensable de supprimer les valeurs extrêmes car elles peuvent influencer exagérément le calcul de la moyenne et de l'écart type utilisés par la suite pour centrer et réduire les distributions. Pour centrer les distributions, la valeur (moyenne ou médiane) la plus proche du sommet des distributions est recherchée. La réduction[157] se fait sur la base de l'écart type annuel. Enfin, deux ajustements sont réalisés : l'un sur la distribution globale, et l'autre sur une demi-distribution. Chacun de ces points fait l'objet d'une sous-partie (D.2. et D.3.).

A. Valeurs extrêmes

Les valeurs extrêmes influencent le calcul de la moyenne et de l'écart type utilisés ensuite pour centrer la distribution. L'étude est confrontée à un paradoxe : supprimer les valeurs extrêmes permet de renforcer la mesure des indicateurs de tendance (moyenne et médiane) mais risque de réduire l'intérêt de la démarche d'ajustement empirique de la loi de distribution. L'étude part en effet du principe que les valeurs extrêmes ne sont pas aberrantes ; elles font partie de la distribution.

Un examen attentif des valeurs extrêmes est mené de manière à ne supprimer que celles qui présentent un caractère aberrant. Cet examen des valeurs extrêmes est le second. Le premier a été conduit dans le cadre de la partie 2. Les données manifestement erronées

[157] Il est important de rappeler que la variable étudiée est la variable résultat mise à l'échelle par l'actif total, donc une variable résultat « déjà » réduite. La réduction par l'écart type permet de standardiser les données, et constitue donc une seconde réduction. Au final, la distribution ajustée est la distribution de la variable [IB/(AT-IB)]/σ.

(erreurs de saisies), ou difficilement interprétables (comme le résultat strictement nul sur plusieurs années) avaient alors été supprimées. La base avait aussi été complétée, notamment dans le cas des données partielles : résultat présent, mais pas le total actif, ou inversement. Il est important de souligner que lors de ce premier examen, les valeurs extrêmes avaient été examinées avec beaucoup d'indulgence dans la mesure où des démarches non paramétriques étaient utilisées pour mesurer les irrégularités. Dans une démarche non paramétrique, seules les données à proximité de l'irrégularité (donc du résultat nul) sont utilisées, et les valeurs situées aux deux extrémités de la distribution n'influencent pas les mesures.

1. Caractéristiques de la distribution

Dans cette partie 3, l'étude s'attache à ajuster la distribution des résultats réduits. L'ajustement de la loi dépend de toutes les valeurs de la distribution, et il sera tout spécialement influencé par les valeurs situées aux extrémités de la distribution. Ces valeurs, malgré leur petit nombre, peuvent influencer considérablement les paramètres de la distribution que sont la moyenne, l'écart type, et dans une moindre mesure la médiane.

Tableau 64 : Rappel des paramètres des distributions annuelles de la variable IB/(AT-IB)

	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**Total**
Nombre	291	296	376	422	616	702	753	798	791	734	698	659	610	7746
Moyenne	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0463	0.0233	0.0014	0.0012	-0.0002	0.0178	
Médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.0350	0.0346	0.0281	0.0214	0.0172	0.0185	0.0259	
Écart type	0.0545	0.0616	0.0544	0.0618	0.0665	0.0791	0.0865	0.2580	0.1185	0.1329	0.1855	0.1375	0.1350	
Max	0.2136	0.2317	0.2905	0.2493	0.4635	0.5866	0.6505	**6.8174**	0.8014	0.5837	**1.9837**	**1.8059**	**1.8092**	
Min	-0.2224	-0.3144	-0.3103	-0.4214	-0.4924	**-0.7097**	**-0.5816**	-0.4654	**-0.8006**	**-0.7983**	**-0.9809**	**-0.5943**	**-0.7301**	
	Les cases en **gras** représentent les années où figurent des valeurs extrêmes.													

2. Règles de détermination des valeurs extrêmes

Comment déterminer qu'une valeur extrême est aberrante statistiquement ? Deux critères de suppression sont retenus. (1) Le cas isolé, qui est identifié grâce à la représentation de la boite à moustaches, et (2) les cas où le dénominateur est manifestement non significatif.

Dans un premier temps, l'étude fixe arbitrairement des bornes au-delà desquelles les résultats sont considérés comme extrêmes. Dans le cadre de notre étude, un premier

examen des cas où le résultat (IB) est égal ou supérieur à 50 % de l'actif total (en valeur absolu) a été mené. Comme l'étude s'intéresse à la variable IB/(AT-IB), lorsque le résultat est supérieur à 50% de l'actif total, alors la variable étudiée est supérieure à 1. Lorsque le résultat est négatif, alors la variable est inférieure à -0,33. Cette première analyse conduit à identifier 6 valeurs positives, et 110 valeurs négatives.

Tableau 65 : Examen des valeurs extrêmes de la distribution

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	
Var > 1	0	0	0	0	0	0	0	1	0	0	3	1	1	**6**
Var < -0,33	0	0	0	1	2	2	3	7	14	28	25	18	10	**110**
Var < -0,5	0	0	0	0	0	2	1	0	5	8	9	4	1	30

En ce qui concerne les valeurs négatives, il apparaît presque chaque année que les sociétés publient un résultat (mis à l'échelle) compris entre -0,33 et -0,5. Ces valeurs ne sont pas isolées, et il ne semble donc pas aberrant statistiquement de conserver ces valeurs. Elles correspondent à la queue épaisse de la distribution à gauche déjà discutée dans le chapitre VII. D'autant plus que cela semble conjoncturel : les années de crise (2000 à 2004) correspondent à un accroissement du nombre des entreprises fortement déficitaires. Les cas où la variable est inférieure à -0,5 sont nettement plus rares. Seules les valeurs hors de l'intervalle]-0,5 ; 1[ont donc été considérées comme extrêmes. Elles font l'objet d'un examen au cas par cas.

3. Analyse systématique de chaque année

Chacune des distributions est triée pour examiner les valeurs extrêmes. Cet examen est détaillé en Annexe 11 : Second examen des valeurs extrêmes au cas par cas. En définitive, 9 données sont supprimées de la base, correspondant à 5 sociétés.

Tableau 66 : Suppression des 9 valeurs extrêmes de la distribution

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	**total**
Var > 1	0	0	0	0	0	0	0	1	0	0	3	1	1	**6**
Var < -0,5	0	0	0	0	0	2	1	0	5	8	9	4	1	**30**
Suppressions	0	0	0	0	0	0	0	1	0	0	5	2	1	**9**

Une société est supprimée sur trois années du fait de variations extrêmes (2002 à 2004), 6 données (dont une seule négative en 2002) ont été retirées du fait d'une situation isolée

de la population, à chaque fois cela s'expliquant par un dénominateur peu significatif. Les nouveaux paramètres des distributions après élimination des 9 données sont présentés (Tableau 67).

Tableau 67 : Caractéristiques de la distribution après suppression des 9 valeurs extrêmes

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Min	Max
données initiales (7746 observations)															
Nombre	291	296	376	422	616	702	753	798	791	734	698	659	610		
Moyenne	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0463	0.0233	0.0014	0.0012	-0.0002	0.0178	0.000	**0.046**
Médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.0350	0.0346	0.0281	0.0214	0.0172	0.0185	0.0259	0.017	0.035
Écart type	0.0545	0.0616	0.0544	0.0618	0.0665	0.0791	0.0865	0.2580	0.1185	0.1329	0.1855	0.1375	0.1350	0.054	**0.258**
Max	0.2136	0.2317	0.2905	0.2493	0.4635	0.5866	0.6505	6.8174	0.8014	0.5837	1.9837	1.8059	1.8092	0.214	6.817
Min	-0.2224	-0.3144	-0.3103	-0.4214	-0.4924	-0.7097	-0.5816	-0.4654	-0.8006	-0.7983	-0.9809	-0.5943	-0.7301	-0.981	-0.222
données corrigées après suppression des 9 données (7737 observations)															
Nombre	291	296	376	422	616	702	753	797	791	734	693	657	609		
Moyenne	0.028	0.0199	0.029	0.0253	0.0321	0.0341	0.0406	**0.0378**	0.0233	0.0014	**-0.004**	**-0.002**	**0.0148**	-0.004	**0.041**
Médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.035	0.0346	0.0281	0.0214	0.0172	0.0185	0.0256	0.017	0.035
Écart type	0.0545	0.0616	0.0544	0.0618	0.0665	0.0791	0.0865	**0.0947**	0.1185	0.1329	0.1305	0.1161	0.1139	0.054	**0.133**
Max	0.2136	0.2317	0.2905	0.2493	0.4635	0.5866	0.6505	0.5495	0.8014	0.5837	0.6431	0.4456	0.7653	0.214	0.801
Min	-0.222	-0.314	-0.31	-0.421	-0.492	-0.71	-0.582	-0.465	-0.801	-0.798	-0.701	-0.594	-0.73	-0.801	-0.222

- Les suppressions n'ont presque pas modifié la médiane.

- Elles ont en revanche conduit à une diminution de la moyenne les 4 années concernées (1999, 2002, 2003 et 2004). La moyenne devient négative en 2002 et 2003 (au lieu de la seule année 2003 auparavant). La distance entre minimum et maximum passe de 0,046 à 0,045. S'il n'y a pas véritablement de modification en valeur, il y a une translation. L'intervalle [0 ; 0,046] devient [-0,004 ; 0,041]. En résumé, les 9 suppressions modifient très sensiblement les moyennes.

- Les écarts types ont été largement modifiés (en 1999 surtout). La distance entre minimum et maximum a diminué (de 0,204 à 0,079), ce qui tend à rendre les écarts types plus homogènes sur la période. L'écart type diminue de 61,27 % alors que 9 valeurs seulement (soit 0,1 % de la base) ont été supprimées.

Tableau 68 : Analyse des différences entre les distributions avant et après suppression des valeurs extrêmes

	1992 à 1998	**1999**		2000 à 2001	**2002**		**2003**		**2004**	
			En %			**En %**		**En %**		**En %**
Écart de N	0	1	0,1%	0	5	0,7%	2	0,3%	1	0,2%
écart de moyenne	0.0	0.0085	**18%**	0.0	0.0056	**467%**	0.0019	**10%**	0.0029	**16%**
écart de médiane	0.0	0.0000	0%	0.0	0.0000	0%	0.0000	0%	0.0003	1%
écart d'ecartype	0.0	0.1633	**63%**	0.0	0.0550	**30%**	0.0214	**16%**	0.0211	**16%**

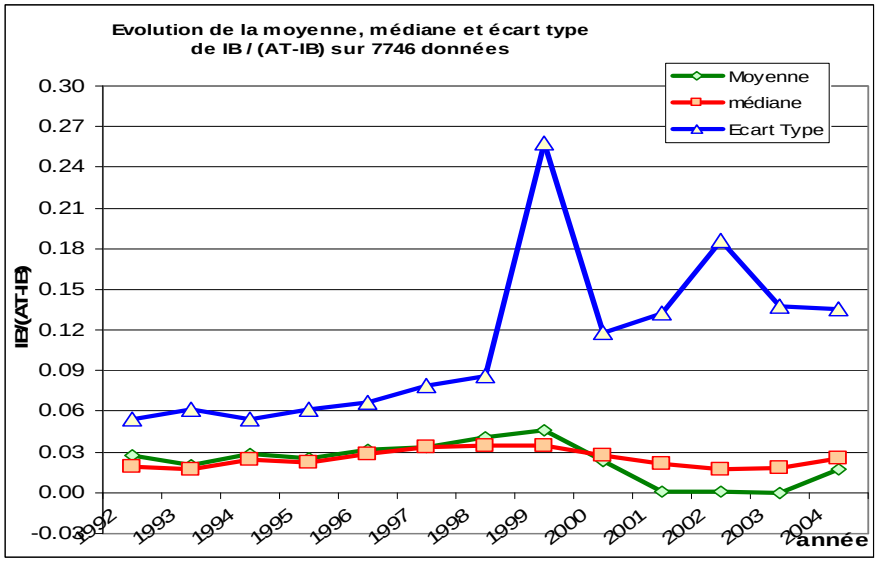
En conclusion, la médiane apparaît comme beaucoup moins sensible aux valeurs extrêmes. Comme cela est décrit plus loin, la médiane apparaît également comme étant plus proche du sommet de la distribution. C'est pourquoi il est préférable de centrer sur la médiane plutôt que sur la moyenne.

B. Réduction

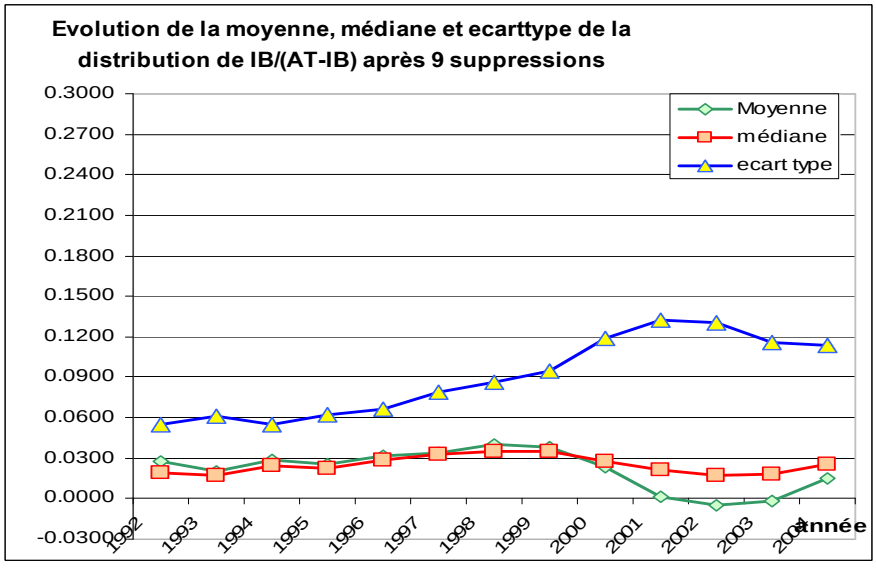
La réduction se fait par l'écart type. Comme le montre la représentation graphique (Figure 99), la suppression des 9 observations extrêmes permet de calculer un écart type relativement stable dans le temps.

Figure 99 : évolution des indicateurs de tendance des distributions avant et après correction

avant correction



après correction

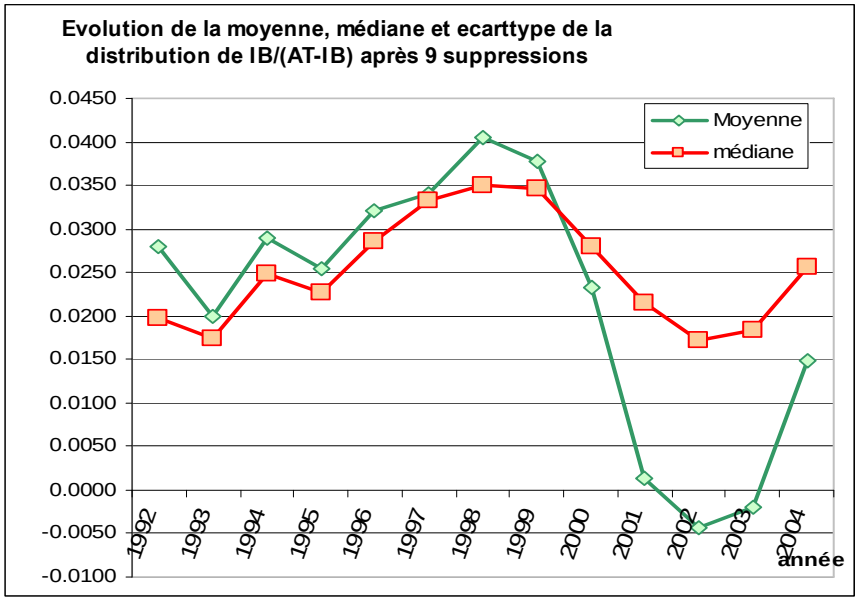


L'écart type évolue sans dent de scie entre 0,06 et 0,14 alors que sans correction, il laisse apparaître des pics de variations allant jusqu'à 500 %. C'est pourquoi il est important de corriger la base des valeurs extrêmes avant la réduction.

C. Centrage

La démarche habituelle consiste à centrer sur la moyenne, celle-ci étant généralement considérée comme le meilleur indicateur de la tendance, et située au sommet de la distribution. À droite et gauche de la moyenne, les pentes de la distribution changent de signe (et d'allure si la distribution n'est pas symétrique). Ce postulat mérite d'être vérifié.

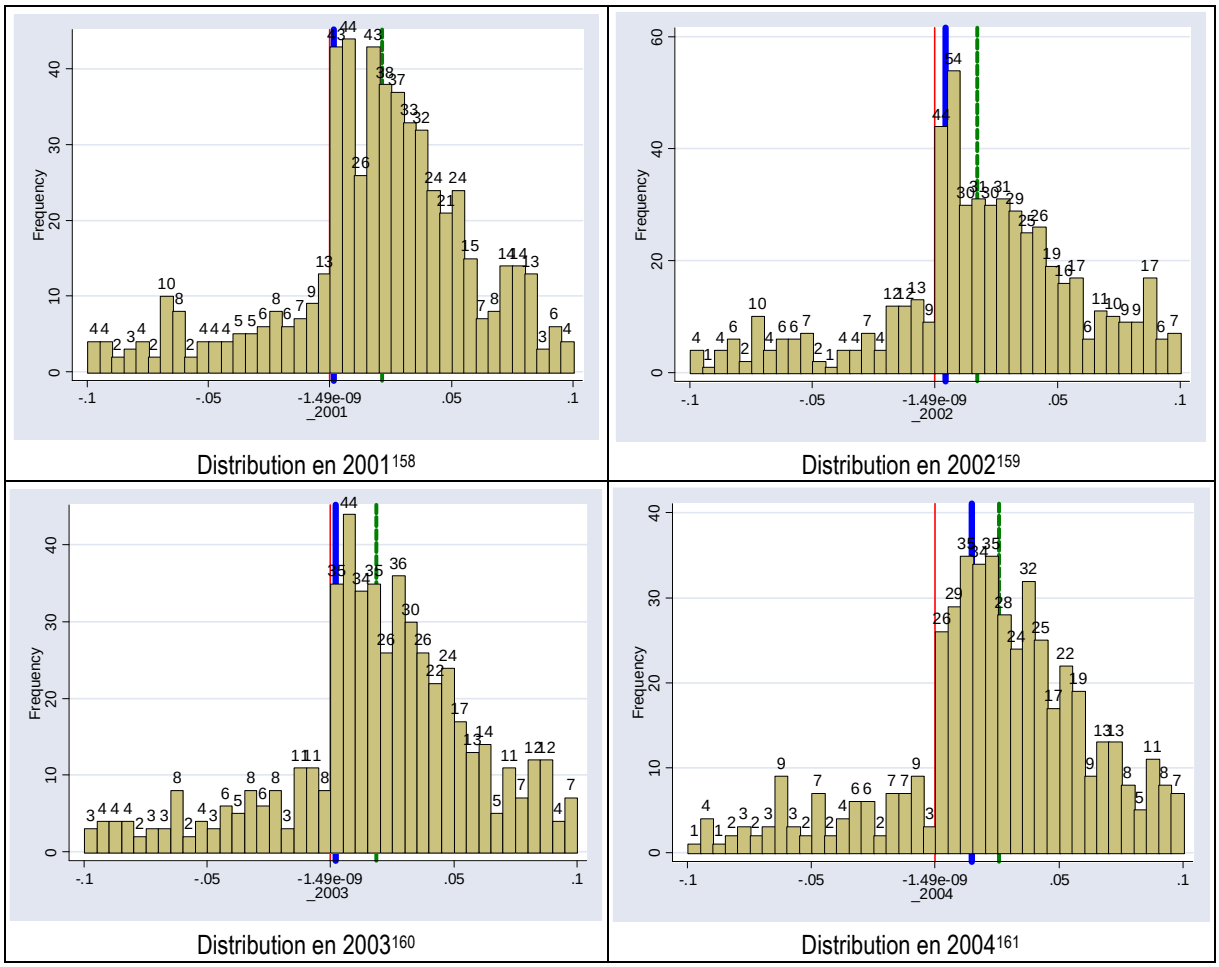
Figure 100 : Comparaison de la moyenne et de la médiane de chacune des distributions annuelles



Il apparaît (Figure 100) que la moyenne, la médiane et le sommet de la distribution se confondent (à très peu de chose près) sur la période 1992 et 2000. Il semble donc indifférent de retenir l'une plutôt que l'autre. Mais entre 2001 et 2004, la médiane et la moyenne s'écartent très sensiblement. Il apparaît alors légitime de se poser la question : quel est, entre ces deux indicateurs, celui qui est situé le plus proche du sommet de la courbe ?

L'observation des distributions durant ces 4 années permet de répondre à cette question. Sur la Figure 101, la moyenne est tracée en épais, et la médiane en pointillé. Le zéro (résultat nul) est indiqué par un trait fin.

Figure 101 : Identification du sommet des distributions entre 2001 et 2004



Distribution en 2001[158] | Distribution en 2002[159]

Distribution en 2003[160] | Distribution en 2004[161]

De 2001 à 2003, la moyenne se confond presque avec le seuil du résultat nul. Il est donc difficile de savoir si le sommet observé en ce point est régulier ou non. En 2001, 2003 et 2004, la distribution apparaît comme étant bimodale. Si la moyenne correspond à un sommet, ce sommet n'est pas seul. Durant ces quatre années, le sommet se situe entre la médiane et la moyenne. Le choix entre médiane et moyenne apparaît donc comme arbitraire.

[158] Fonction sous Stata : histogram _2001 if _2001>-0.1 & _2001<0.1, width(0.005) start(-0.100000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.00135, lcolor(blue) lwidth(vthick)) xline(0.0214, lcolor(green) lwidth(thick) lpattern(dash)).

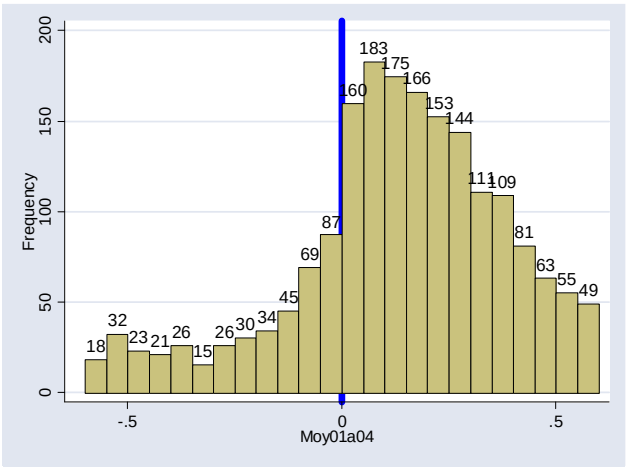
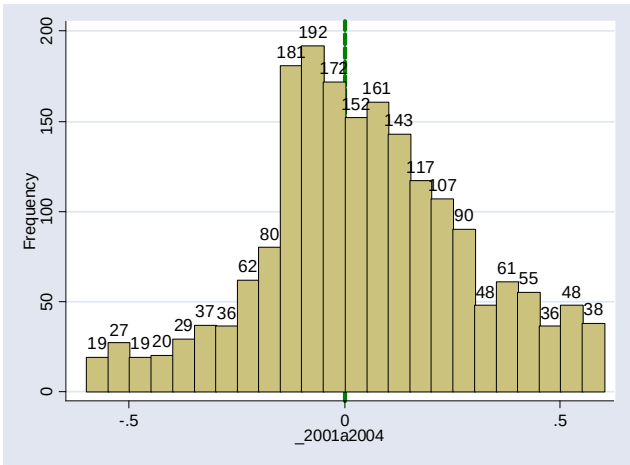
[159] Fonction sous Stata : histogram _2002 if _2002>-0.1 & _2002<0.1, width(0.005) start(-0.100000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.00436, lcolor(blue) lwidth(vthick)) xline(0.0172, lcolor(green) lwidth(thick) lpattern(dash)).

[160] Fonction sous Stata : histogram _2003 if _2003>-0.1 & _2003<0.1, width(0.005) start(-0.100000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.00206, lcolor(blue) lwidth(vthick)) xline(0.0185, lcolor(green) lwidth(thick) lpattern(dash)).

[161] Fonction sous Stata : histogram _2004 if _2004>-0.1 & _2004<0.1, width(0.005) start(-0.100000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.0148, lcolor(blue) lwidth(vthick)) xline(0.0256, lcolor(green) lwidth(thick) lpattern(dash)).

Les distributions des résultats (mis à l'échelle) agrégés de 2001 à 2004, centrés et réduits sur la médiane et la moyenne, sont comparées.

Figure 102 : Comparaison des sommets des distributions agrégées de 2001 à 2004 centrées sur moyenne et médiane

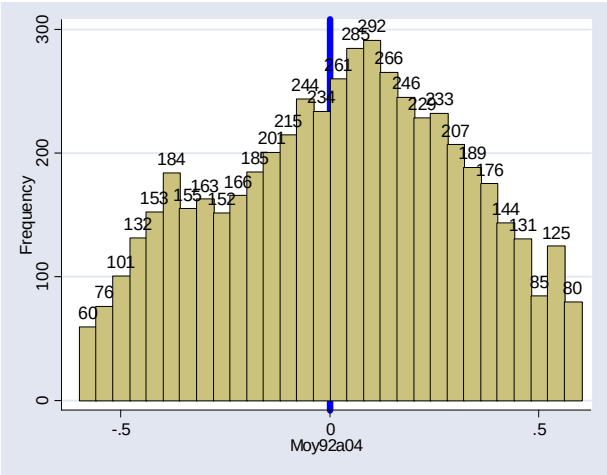
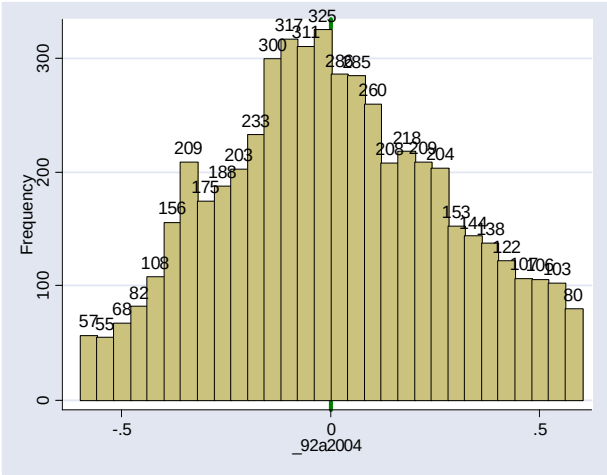
Distribution centrée sur la moyenne[162]		Distribution centrée sur la médiane[163]	
			
N	2693	N	2693
mean	0	mean	-0.1481087
p50	0.1510249	p50	0
sd	0.9994426	sd	1

Il apparaît manifestement (Figure 102) que le sommet est situé entre la moyenne et la médiane. Les distributions (centrées réduites sur médiane et moyenne) de toutes les observations agrégées (de 1992 à 2004) sont également tracées (Figure 103). La médiane apparaît alors plus proche du sommet.

[162] Fonction sous Stata : histogram Moy01a04 if Moy01a04>-0.6 & Moy01a04<0.6, width(0.05) start(-0.600000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(blue) lwidth(vthick)).

[163] Fonction sous Stata : histogram _2001a2004 if _2001a2004>-0.6 & _2001a2004<0.6, width(0.05) start(-0.600000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(green) lwidth(thick) lpattern(dash)).

Figure 103 : Comparaison des sommets des distributions agrégées de 1992 à 2004 centrées sur moyenne et médiane

Distribution centrée sur la moyenne[164]		Distribution centrée sur la médiane[165]	
			
N	7737	N	7737
Mean	0	mean	-0.02725
p50	0.0590676	p50	0
Sd	0.9992241	sd	1.00395

En conclusion, l'étude retient la médiane pour centrer les distributions, car (1) elle est plus proche du sommet de la distribution, (2) elle est moins sensible aux valeurs extrêmes, et (3) elle ne se confond jamais avec le seuil du résultat nul.

D. Ajustement

L'ajustement est réalisé sur la distribution de la variable résultat annuel (IB) mis à l'échelle par l'actif total (AT-IB), centré sur la médiane annuelle, et réduit par l'écart type annuel.

1. Paramètres de la distribution étudiée

Les tableaux (Tableau 69 et Tableau 70) présentent les paramètres des distributions avant et après réduction et centrage.

[164] Fonction sous Stata : histogram Moy92a04 if Moy92a04>-0.6 & Moy92a04<0.6, width(0.04) start(-0.600000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(blue) lwidth(vthick)).

[165] Fonction sous Stata : histogram _92a2004 if _92a2004>-0.6 & _92a2004<0.6, width(0.04) start(-0.600000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(green) lwidth(thick) lpattern(dash)).

Tableau 69 : Paramètres des distributions annuelles « brutes » (résultat mis à l'échelle par (AT-IB) non centré-réduit)

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Nombre	291	296	376	422	616	702	753	797	791	734	693	657	609
Moyenne	0.0280	0.0199	0.0290	0.0253	0.0321	0.0341	0.0406	0.0378	0.0233	0.0014	-0.0044	-0.0021	0.0148
médiane	0.0197	0.0173	0.0249	0.0227	0.0285	0.0333	0.0350	0.0346	0.0281	0.0214	0.0172	0.0185	0.0256
Écart type	0.0545	0.0616	0.0544	0.0618	0.0665	0.0791	0.0865	0.0947	0.1185	0.1329	0.1305	0.1161	0.1139
Max	0.2136	0.2317	0.2905	0.2493	0.4635	0.5866	0.6505	0.5495	0.8014	0.5837	0.6431	0.4456	0.7653
min	-0.2224	-0.314	-0.310	-0.421	-0.492	-0.709	-0.581	-0.465	-0.800	-0.798	-0.700	-0.594	-0.730
IQR	0.0474	0.0516	0.0500	0.0513	0.0543	0.0544	0.0631	0.0629	0.0556	0.0597	0.0634	0.0612	0.0544
2IQRn-1/3	0.0143	0.0155	0.0138	0.0137	0.0128	0.0122	0.0139	0.0136	0.0120	0.0132	0.0143	0.0141	0.0128

Tableau 70 : Paramètres des distributions annuelles centrées-réduites

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Nombre	291	296	376	422	616	702	753	797	791	734	693	657	609
Moyenne	0.1521	0.0418	0.0749	0.0427	0.0547	0.0101	0.0649	0.0340	-0.0398	-0.1511	-0.1649	-0.1767	-0.0946
médiane	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
écart type	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Max	3.5593	3.4792	4.8778	3.6654	6.5392	6.9963	7.1151	5.4355	6.5245	4.2320	4.7956	3.6778	6.4950
min	-4.4446	-5.3851	-6.1555	-7.1837	-7.8298	-9.3940	-7.1262	-5.2778	-6.9917	-6.1698	-5.4989	-5.2765	-6.6359
IQR	0.8704	0.8375	0.9179	0.8301	0.8161	0.6878	0.7288	0.6639	0.4695	0.4495	0.4855	0.5270	0.4774
2IQRn-1/3	0.2627	0.2513	0.2543	0.2213	0.1918	0.1548	0.1602	0.1432	0.1015	0.0997	0.1097	0.1212	0.1127

Enfin, le Tableau 71 présente les paramètres de la distribution agrégée qui permet d'ajuster la loi de distribution.

Tableau 71 : Paramètres de la distribution des résultats centrés-réduits agrégés sur 13 ans

Agrégation de 1992 à 2004	Minimum	Maximum	Moyenne	Mode	Median	Std. Deviation	Skewness	Kurtosis
Nombre	-9,394	7,1151	-0,02725	0,00	0	1,0039	-0,9465	13,290

Avec un Skewness de -0,9465, la distribution agrégée servant à l'ajustement est dissymétrique (plus étalée sur la gauche). Avec un Kurtosis de 13,2901, elle est également très pointue. Comme cela a été expliqué dans le chapitre VII, aucune loi ne permet de traduire ces deux caractéristiques. Deux ajustements sont réalisés. Le premier sur la distribution globale, et le second sur chaque moitié de distribution.

2. Ajustement de la distribution globale (contrainte de symétrie)

a. Démarche

Les trois meilleurs ajustements proposés par le logiciel Best Fit sont reproduits dans le Tableau 72. La loi logistique[166] apparaît comme la moins mauvaise. Les tests statistiques laissent cependant apparaître que l'ajustement est très peu significatif. La probabilité de se tromper en considérant que la loi logistique est la loi de distribution des résultats annuels des entreprises françaises est supérieure à 99 %.

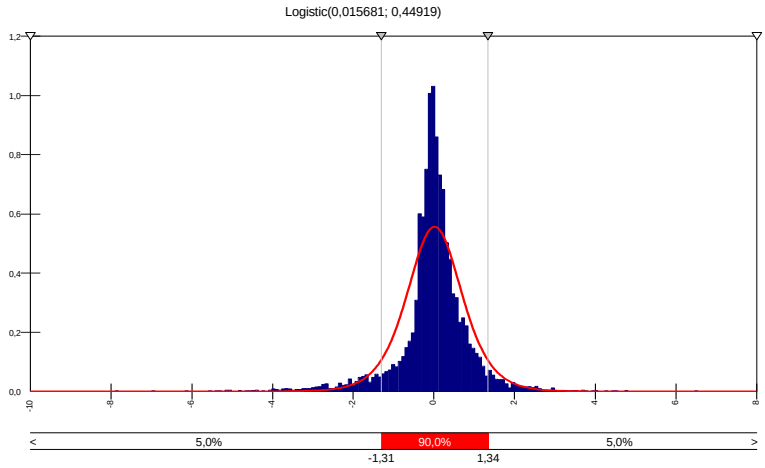
Tableau 72 : Trois ajustements de la distribution globale

	Distribution observée	Loi ajustée		
Function		**=RiskLogistic(0,015681; 0,44919)**	**=RiskStudent(6)**	**=RiskNormal(-0,027249; 1,0039)**
Parameter 1		0,0156808695968394	6	-0,0272485854077822
Parameter 2		0,449186294647937	N/A	1,00394757040528
Minimum	-9,3940	-Infinity	-Infinity	-Infinity
Maximum	7,1151	+Infinity	+Infinity	+Infinity
Mean	-0,027249	0,015681	0,0000	-0,027249
Mode	0,0000	0,015681	0,0000	-0,027249
Median	0,0000	0,015681	0.0000	-0,027249
Std. Deviation	1,0039	0,81473	1,2247	1,0039
Skewness	-0,9465	0,0000	0,0000	0,0000
Kurtosis	13,2901	4,2000	6,0000	3,0000
A-D Test Value		**132,9**	**388,0**	**349,4**
P Value		< 0,005	N/A	< 0,005
Rank		1	3	2
K-S Test Value		**0,09746**	**0,1669**	**0,1702**
P Value		< 0,01	N/A	< 0,01
Rank		1	2	3
Chi-Sq Test Value		**1901**	**3873**	**3767**
P Value		0,0000	0,0000	0,0000
Rank		1	3	2

Il n'en demeure pas moins que, même si l'ajustement peut paraître très approximatif, la loi logistique est celle qui traduit le mieux les données selon les trois critères (test du Chi², test de K-S, test de A-D). Elle est notamment plus pointue que la loi normale.

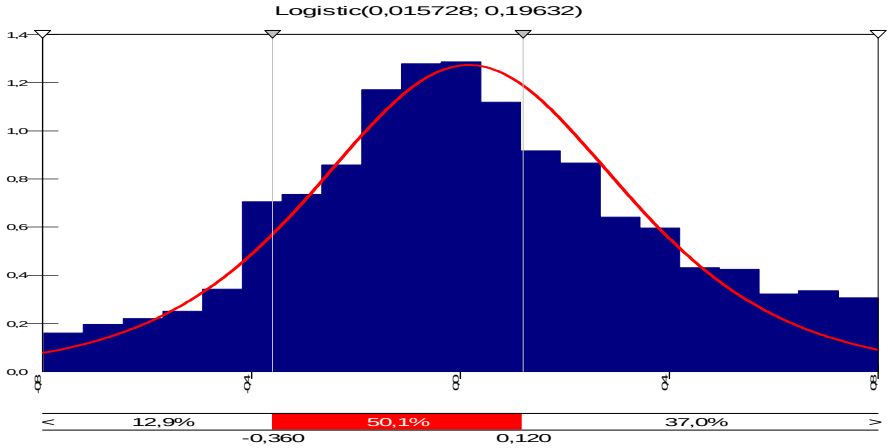
[166] Les paramètres de la fonction Logistique dans le logiciel Best Fit sont les suivants : Fonction RISKLogistic(α, β) où α est un paramètre continu, et β un « paramètre d'échelle » continu supérieur strictement à zéro. $-\omega \leq x \leq +\omega$; Moyenne = α ; Variance = $\pi^2 \beta^2 / 3$; Mode = α ; Skewness = 0 ; Kurtosis = 4,2 ; Fonction de densité $f(x) = \text{sech}^2 (1/2((x-\alpha)/ \beta)) / 4\beta$; Fonction cumulative $F(x) = 1+ \tanh(1/2((x-\alpha)/\beta))/2$ avec « sech » la fonction secante hyperbolique, et « tanh » la fonction tangeante hyperbolique.

Figure 104 : Représentation graphique de l'ajustement de la distribution par une loi logistique



Graphiquement, il apparaît que la loi Logistique ajuste mal les queues de distribution. Afin d'améliorer l'ajustement, la démarche a été poursuivie en ne prenant que la partie centrale de la distribution. La Figure 105 présente la distribution sur l'intervalle]-0,8 ; 0,8[identifié par tâtonnement comme permettant le meilleur ajustement.

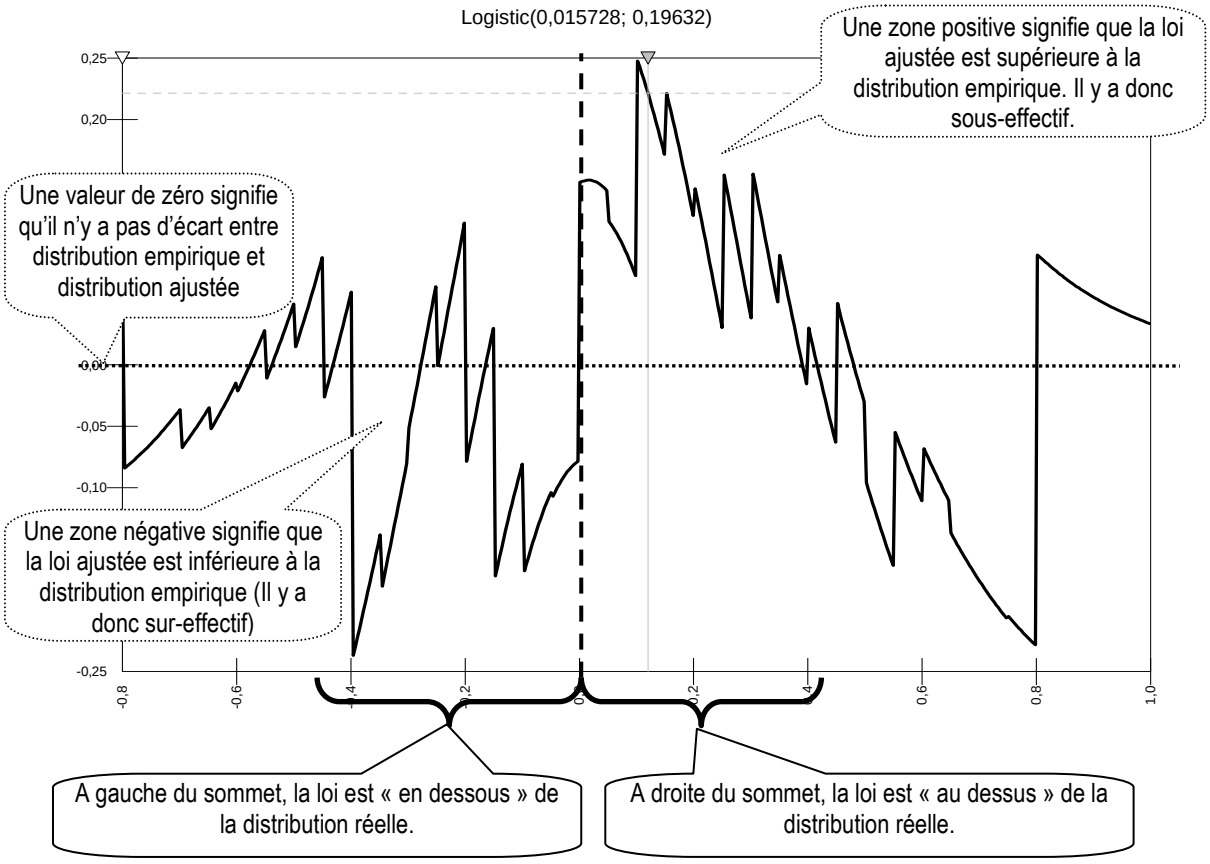
Figure 105 : Ajustement et représentation graphique de l'ajustement sur l'intervalle]-0,8 ; 0,8[



	Distribution observée	Loi ajustée	
Function		**"=RiskLogistic(0,015728; 0,19632)"**	**"=RiskNormal(0,02248; 0,34123)"**
Parameter 1		1,57E-02	2,25E-02
Parameter 2		0,196323	0,341232
Minimum	-0,7981	"=-Infinity"	"=-Infinity"
Maximum	0,79965	"=+Infinity"	"=+Infinity"
Mean	0,02248	0,015728	0,02248
Mode	0,0000 [est]	0,015728	0,02248
Median	-0,00011	0,015728	0,02248
Std. Deviation	0,34123	0,35609	0,34123
Skewness	0,1162	0	0
Kurtosis	2,6494	4,2	3
A-D Test Value		**11,03**	**9,663**
P Value		< 0,005	< 0,005
Rank		2	1
K-S Test Value		**0,0267**	**0,03156**
P Value		< 0,01	< 0,01
Rank		**1**	**2**
Chi-Sq Test Value		**569,9**	**250,4**
P Value		0	0
Rank		2	1

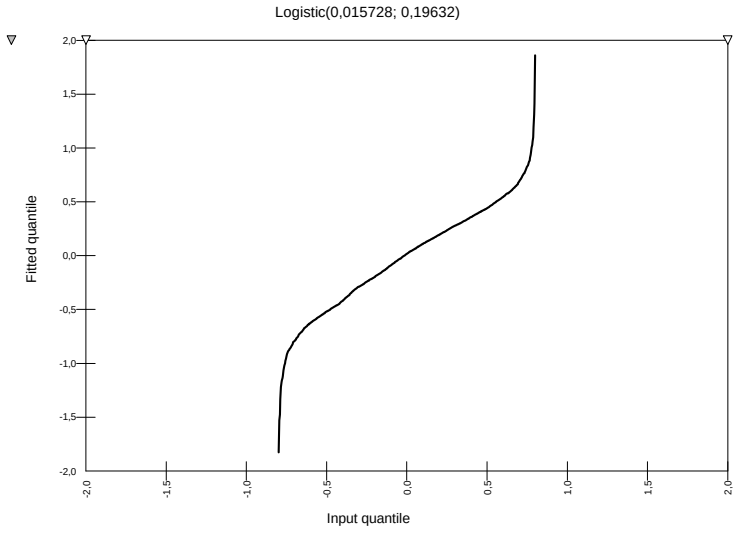
La distance du test de K-S est alors de 0,02670 (Figure 105) ce qui est une nette amélioration par rapport à la distance mesurée sur l'ajustement de la distribution totale (0,09746 lu dans le Tableau 72). Cependant, cet ajustement a deux défauts : l'observation visuelle de la distribution suffit à révéler clairement que la loi logistique ne rend pas compte de la dissymétrie de la distribution empirique. Par ailleurs, cet ajustement ne tient pas compte des queues de distribution (hors de l'intervalle]-0,8 ; 0,8[).

Figure 106 : Distribution des écarts entre distribution empirique et loi ajustée sur l'intervalle]-0,8 ; 0,8[



Les écarts entre distribution empirique et distribution ajustée sont représentés graphiquement sur la Figure 106. On observe graphiquement que les écarts ne sont pas également disposés sur la distribution. À droite du sommet, la distribution est « en dessous » de la loi (elle est plus « creusée ») alors qu'à gauche du sommet, le phénomène inverse se produit. En effet, la distribution est asymétrique, ce que la loi Logistique n'est pas à même de représenter.

Figure 107 : « Q-Q plot » de l'ajustement Logistique



La Figure 107 présente le graphe « Q-Q plot » de l'ajustement Logistique. C'est une représentation graphique de la qualité de l'ajustement. La partie centrale de l'ajustement est presque linéaire ce qui est le signe que l'ajustement est bon. Les extrémités de la courbe s'écartent de la diagonale, ce qui est le signe que l'ajustement n'est pas bon pour représenter les queues de distribution.

b. Synthèse de la démarche d'ajustement de la fonction globale

En conclusion, la partie centrale de la distribution peut être raisonnablement ajustée par une loi logistique. Cet ajustement est le meilleur obtenu empiriquement. Il permet de mesurer les irrégularités de distribution en s'affranchissant des limites inhérentes aux méthodes non paramétriques. Cependant, la loi de distribution Logistique est symétrique, et ajuste très mal les queues de distribution. C'est pourquoi dans un second temps, un deuxième ajustement est réalisé sur des demi-distributions partielles permettant de tenir compte de l'allure non symétrique des distributions observées.

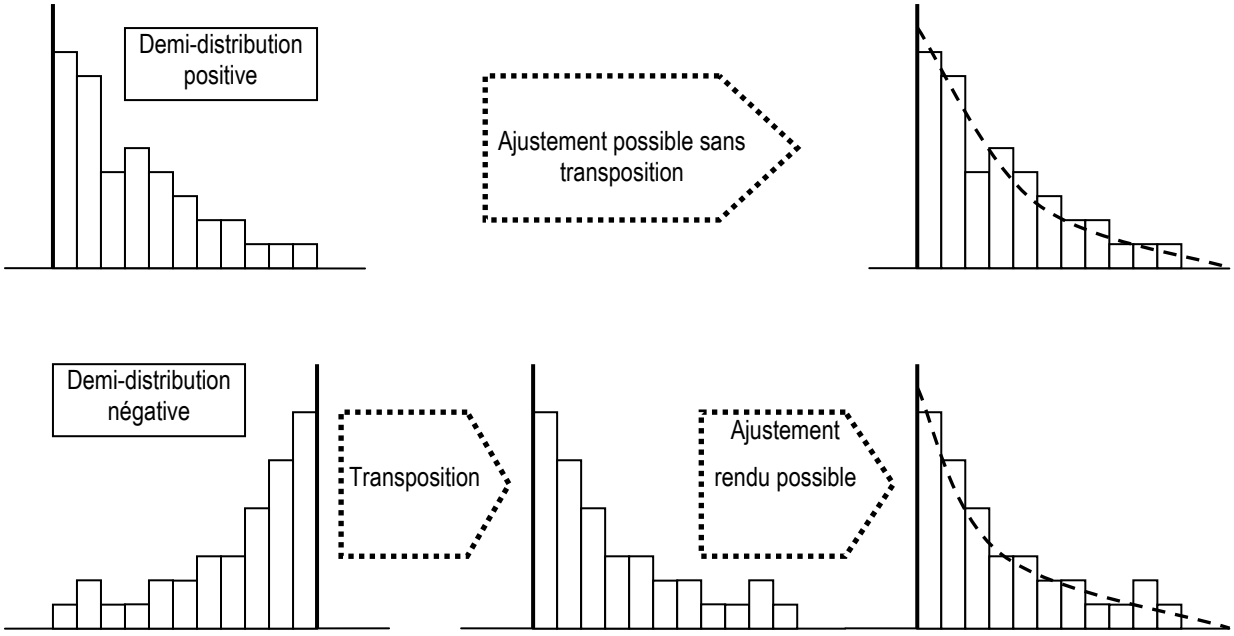
3. Ajustement de la demi-distribution (pas de contrainte de symétrie)

a. Démarche

a.1. Division en deux courbes

Pour résoudre le problème de la dissymétrie, la distribution a été « découpée » en deux : les valeurs positives d'une part et les valeurs négatives d'autre part. Chaque demi-distribution comprend 3873 données[167]. La partie gauche de la courbe (valeurs négatives) est ensuite transposée (en multipliant par -1) de manière à ce que les deux demi-distributions soient de valeurs positives. En effet, les lois monotones testées ne sont définies que sur l'intervalle [0 ; +∞[. La Figure 108 illustre cette transformation.

Figure 108 : Transposition de la demi-distribution des valeurs négatives



[167] La distribution étant centrée sur la médiane, le découpage effectué par rapport au sommet conduit à diviser la population (de taille N) en deux sous-population de même taille (N/2).

a.2. Jonction des deux courbes

Le problème des données strictement nulles

Les rares (7 au total) valeurs strictement nulles peuvent poser problème. Si elles sont laissées telles quelles, elles ne sont pas traitées par le logiciel car la plupart des lois testées ne sont valables que pour des données strictement positives. Il y a perte d'informations. Or ces données sont d'autant plus importantes qu'elles constituent le point de jonction des deux fonctions. Ces valeurs ont été remplacées par des valeurs extrêmement faibles (0,00000000001). Ainsi, les données strictement nulles ont été intégrées dans les deux distributions avec une valeur inférieure à la précision utilisée dans la base (9 chiffres après la virgule). Les lois ont été ajustées avec et sans modification de ces données strictement nulles, et les résultats sont quasiment identiques.

Tableau 73 : Nombre d'observations strictement égales à zéro

Année	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Total
Nombre de données strictement égales à zéro	1	0	0	0	0	0	1	1	1	0	1	1	1	7

La jonction des deux lois

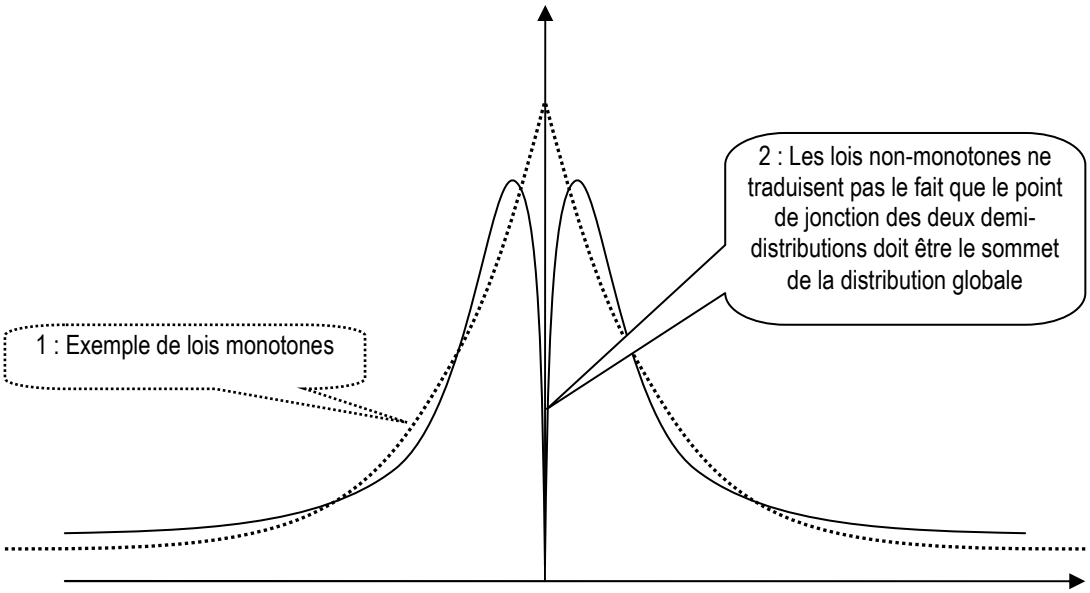
Comme la partie droite et la partie gauche sont ajustées indépendamment, les deux lois ne se rejoignent pas exactement en leurs sommets. Ce problème de jonction inhérent à la démarche d'ajustement par découpage est préoccupant. Il n'y a pas à notre connaissance d'outil mathématique permettant de résoudre ce problème. Sa conséquence est de rendre les estimations proches du sommet moins fiables. Les comparaisons de mesures effectuées par les deux ajustements (global et partiel) présentés dans le chapitre IX montrent cependant que l'impact de ce problème est manifestement négligeable car les irrégularités apparaissent toujours sur la partie gauche de la distribution. Aucune estimation n'est ainsi faite au-delà du sommet. L'ajustement de la partie droite est cependant réalisé tout de même. Il permet de confirmer que l'allure à droite est similaire à l'allure à gauche, ce qui renforce les résultats.

a.3. La nature des lois testées

Les deux demi-distributions ont été traitées par le logiciel Best Fit qui a classé les lois d'ajustement. Toutes les lois disponibles sur le logiciel sont testées. Les résultats ne sont présentés que pour la dizaine de lois les mieux ajustées. Plusieurs de ces lois (parfois bien classées) sont non-monotones (Lognormale ou Inverse-Gaussienne). Or la

distribution ayant été scindée en son milieu, ce point doit correspondre au sommet de la distribution.

Figure 109 : Les lois monotones sont privilégiées dans le cadre des ajustements partiels



Une loi non monotone (ligne continue sur la Figure 109) connaît une phase ascendante, puis descendante. Autrement dit, si l'on ajuste chaque demi distribution par une loi non monotone, la jonction des deux lois ne se fait pas au sommet de la distribution (propos n°2 sur la figure). A contrario, les lois monotones peuvent se rejoindre en leur sommet (propos n°1 sur la figure). Les lois monotones (Exponentielle ou Pareto) ont donc été privilégiées par rapport aux lois non monotones.

b. Ajustement des données positives (partie droite de la courbe)

Le Tableau 74 reproduit les principaux ajustements proposés par le logiciel. La loi Pareto2 est la loi qui ajuste au mieux la distribution. Les lois Inverse-Gaussienne et LogNormale2 sont également bien classées, mais ce ne sont pas des lois monotones. La loi Exponentielle est moins bien classée que la loi Pareto2 sur deux des trois critères. Elle n'est meilleure que pour le test A-D qui accorde surtout de l'importance aux queues de distribution. Or l'étude s'intéresse aux manipulations autour du seuil résultat nul qui se situent toujours dans la partie centrale de la distribution. Les tests du Chi² et de K-S sont donc privilégiés.

Tableau 74 : Ajustement de la demi-distribution « droite » (valeurs positives)

	Lois monotones			Lois NON monotones						
	Pareto2	**Expon**	Pareto	**InvGauss**	**Lognorm2**	LogLogistic	Pearson5	ExtValue	Logistic	Normal
Shift	1E-11	-1,47E-04	-	-5,91E-02	-2,55E-02	-	-0,124155	-	-	-
Parameter 1	3,0816711	0,5691277	4,15E-02	0,62819452	-1,04376862	-7,19E-03	1,9562861	0,32101168	0,45418243	0,5691277
Parameter 2	6,41851788	-	1E-11	0,47084671	1,07314279	0,33742681	0,73852533	0,36181223	0,30158729	0,68889616
Parameter 3	-	-	-	-	-	1,46104276	-	-	-	-
Mean	0,56873	0,56898	-	0,56913	0,60081	0,85989	0,64813	0,52985	0,45418	0,56913
Mode	0.0000	-0,00014699	0.0000	0,089146	0,085843	0,10004	0,12566	0,32101	0,45418	0,56913
Median	0,35143	0,39434	0,00017625	0,32496	0,32665	0,33024	0,32756	0,45362	0,45418	0,56913
Std. Deviation	0,68546	0,56913	-	0,72561	0,92116	-	-	0,46404	0,54702	0,6889
Skewness	3,6011	2	-	3,4652	7,5945	-	-	1,1395	0	0
Kurtosis	32,975	9	-	23,0127	190,4685	-	-	5,4	4,2	3
A-D Test Value	#NOM?	12,91	#NOM?	7,128	7,754	14,13	13,53	96,7	152,1	#NOM?
P Value	-	< 0,01	-	-	-	-	-	< 0,01	< 0,005	< 0,005
Rank	14	**3**	13	**1**	**2**	5	4	6	7	12
K-S Test Value	0,01088	0,04376	0,5571	0,03113	0,03131	0,03049	0,04165	0,1067	0,1815	0,2044
P Value	-	< 0,01	-	-	-	-	-	< 0,01	< 0,01	< 0,01
Rank	**1**	**6**	13	**3**	**4**	**2**	5	7	8	10
Chi-Sq Test Value	52,21	107,9	52753	119	125,1	248,9	299	1394	2720	3820
P Value	0,3882	3,90E-06	0	1,50E-07	2,27E-08	0	0	0	0	0
Rank	**1**	**2**	15	**3**	**4**	**5**	6	7	9	10

La distance de K-S est deux fois plus faible que celle mesurée pour l'ajustement global par la loi Logistique sur la seule partie centrale de la distribution (0,01088 contre 0,0226). Autrement dit, l'ajustement partiel fournit de bien meilleurs résultats que l'ajustement global. La P-value du test du Chi^2 suggère qu'il y a près de 61 % de chances de se tromper en considérant que la distribution empirique est issue d'une loi Pareto2. Ce risque est important en valeur absolue, mais en valeur relative, il peut être considéré comme très correct car l'objectif est moins d'ajuster une loi de distribution parfaite, que d'ajuster la meilleure loi possible. Les avantages de la méthode paramétrique sont tels que même un ajustement imparfait vaut mieux que des estimations non paramétriques dont les nombreuses limites ont été étudiées en partie 2.

c. Ajustement des données négatives (partie gauche de la courbe)

En ce qui concerne les données négatives, force est de constater que l'allure de la distribution est moins régulière que pour la partie droite. Les tests d'ajustement classent la loi Inverse-Gaussienne comme étant le meilleur ajusteur, suivie de la loi Lognorm2. Les

lois monotones (Pareto2 et Exponentielle) n'apparaissent qu'en troisième, quatrième ou cinquième position selon les tests.

Tableau 75 : Ajustement de la demi-distribution « gauche » (valeurs négatives)

	Lois monotones		Lois non monotones					
	Pareto2	Expon	**InvGauss**	Lognorm2	Pearson5	ExtValue	Rayleigh	Logistic
Shift	1E-11	-1,61E-04	-3,40E-02	-8,84E-03	-6,43E-02	-	-0,60894744	-
Parameter 1	0,87295815	0,6235756	0,65755788	-1,21950415	1,3175499	0,30930161	1,08331146	0,44071136
Parameter 2	2,33213377	-	0,28406311	1,28274614	0,33867774	0,42709466	-	0,37947201
Mean	0,65531	0,62341	0,62358	0,66363	1,0022	0,55582	0,74878	0,44071
Mode	0.0000	-0,00016105	0,058819	0,048142	0,081817	0,3093	0,47436	0,44071
Median	0,30214	0,43207	0,28064	0,28653	0,27326	0,46584	0,66655	0,44071
Std. Deviation	1,7365	0,62358	1,0004	1,3754	-	0,54777	0,70972	0,68829
Skewness	-	2	4,5644	14,692	-	1,1395	0,6311	0
Kurtosis	-	9	37,7225	1077,9126	-	5,4	3,2451	4,2
A-D Test Value	#NOM?	154	3,541	3,663	6,961	245,6	#NOM?	301,5
P Value	-	< 0,01	-	-	-	< 0,01	-	< 0,005
Rank	13	**4**	**1**	**2**	**3**	5	14	6
K-S Test Value	**0,04514**	0,1548	0,02684	0,0307	0,02734	0,1828	0,282	0,2384
P Value	-	< 0,01	-	-	-	< 0,01	-	< 0,01
Rank	**4**	**5**	**1**	**3**	**2**	6	9	7
Chi-Sq Test Value	188,3	770	158,4	165,8	235,3	3164	5147	5226
P Value	0	0	0	0	0	0	0	0
Rank	**3**	**5**	**1**	**2**	**4**	6	7	8

La fonction Pareto2 est retenue comme loi d'ajustement pour trois raisons :

1. Elle est monotone.
2. Elle apparaît mieux classée que la loi exponentielle sur les deux critères les plus importants (test K-S et test Chi²).
3. Elle a été retenue pour ajuster la partie droite de la distribution, et cette homogénéité est un gage de cohérence.

En définitive, la loi Pareto2[168] est le meilleur ajusteur de la demi-distribution coupée sur la médiane. Cette allure est, aux paramètres près, la même sur les deux demi-distributions.

[168] Les paramètres de la fonction Pareto (seconde forme) dans le logiciel Best Fit sont les suivants : Fonction RISKPareto2(b,q) où b est un « paramètre d'échelle » continu supérieur strictement à zéro, et q est un « paramètre de forme » continu supérieur strictement à zéro. $0 \leq x \leq +\omega$; Moyenne = $b / (q-1)$; Variance = $b^2q / [(q-1)^2 (q-2)]$; Mode = 0 ; Skewness = $2 \times ((q+1)/(q-3)) \times \sqrt{[(q-2)/q]}$ avec $q>3$; Kurtosis = $[3 \times (q-2) \times (3q^2 +q +2)] / [q \times (q-3) \times (q-4)]$ avec $q>4$; Fonction de densité $f(x) = qbq / (x+b)q+1$; Fonction cumulative $F(x) = 1- (bq)/(x+b)q$.

Le fait que l'ajustement soit meilleur sur la partie droite peut s'interpréter par le fait que la partie gauche contient l'irrégularité au seuil du résultat nul. La zone inquiétante n'a pas été retirée puisque le centrage et la réduction tendent à lisser cette irrégularité. Cependant, l'irrégularité n'a pas disparu, et se traduit par le fait que la distribution est plus lisse à droite qu'à gauche[169]. L'utilisation d'une même loi à droite et à gauche de la distribution est un gage de cohérence.

Pour contrôler les résultats, un critère de validation externe est recherché. Pour cela, l'ajustement de la distribution des résultats des entreprises britanniques et allemandes a été réalisé. Si l'allure de la distribution (la loi d'ajustement) est la même dans ces trois pays, cela renforce l'hypothèse que la loi de distribution des résultats est universelle. Un résultat opposé tendrait à fragiliser la démarche, sans nécessairement la rejeter totalement.

4. Extension de la méthode à d'autres populations : une étude internationale

La même démarche d'ajustement de lois de distribution est appliquée aux échantillons britannique et allemand de manière à comparer les lois qui ajustent le mieux les distributions.

a. Composition des échantillons internationaux

Les données relatives aux entreprises allemandes et britanniques sont extraites de la base Global sur la période 1995 à 2005[170]. Du fait du grand nombre de données manquantes, l'année 2005 n'est pas retenue. Aucune recherche des valeurs aberrantes n'est menée, comme cela a été fait pour les entreprises françaises, mais deux observations extrêmes sont cependant identifiées et supprimées dans l'échantillon britannique et une valeur extrême dans l'échantillon allemand.

[169] Burgstahler et Dichev (1997) estiment que la moitié droite de la distribution est présumée être moins manipulée que la moitié gauche.

[170] Au moment de la consultation de la base, il n'a pas été possible de remonter au-delà de 10 années. L'échantillon des entreprises allemandes et britanniques est donc prélevé sur une période sensiblement plus limitée que l'échantillon d'entreprises françaises (10 ans au lieu de 13).

a.1. Analyse des données

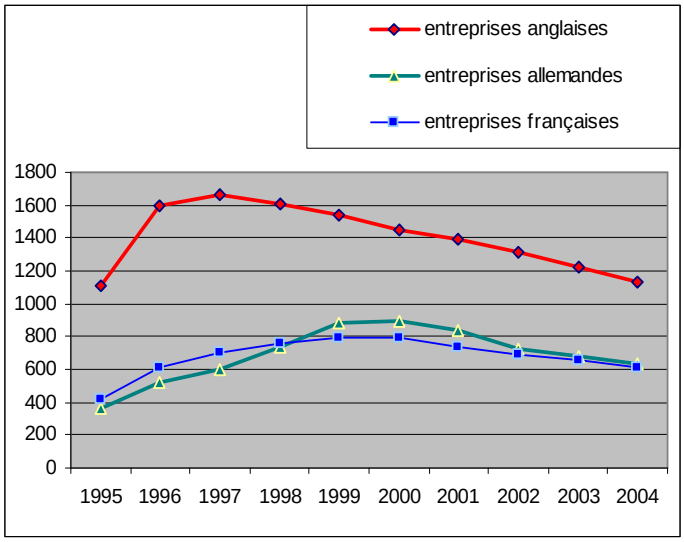
Tableau 76 : Caractéristiques comparées des résultats mis à l'échelle des entreprises allemandes, britanniques et françaises

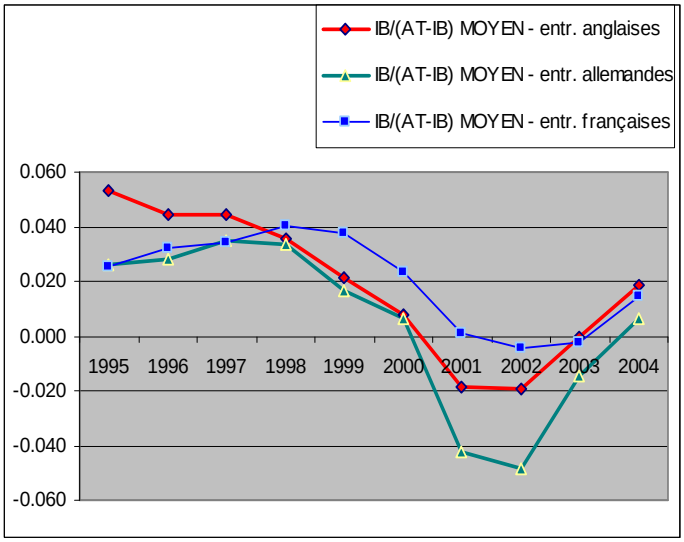
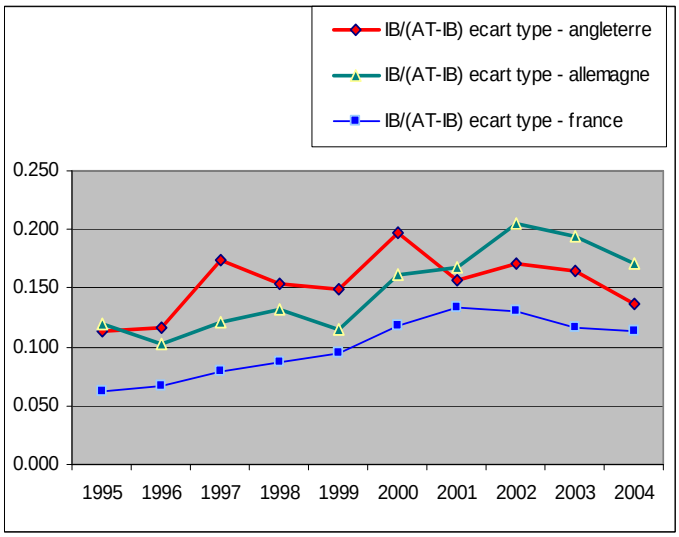
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	total
entreprises britanniques											
Nombre	1112	1595	1664	1609	1539	1448	1398	1311	1220	1132	14 028
Moyenne	0.054	0.045	0.044	0.036	0.021	0.008	-0.019	-0.019	0.000	0.018	
Écart type	0.113	0.116	0.174	0.154	0.149	0.198	0.157	0.171	0.165	0.137	
entreprises allemandes											
Nombre	365	526	598	737	886	899	837	721	679	631	6 879
Moyenne	0.026	0.028	0.035	0.033	0.017	0.006	-0.042	-0.049	-0.015	0.007	
Écart type	0.119	0.103	0.122	0.133	0.114	0.161	0.167	0.205	0.194	0.171	
entreprises françaises											
Nombre	422	616	702	753	797	791	734	693	657	609	6 774
Moyenne	0.025	0.032	0.034	0.041	0.038	0.023	0.001	-0.004	-0.002	0.015	
Écart type	0.062	0.067	0.079	0.087	0.095	0.119	0.133	0.131	0.116	0.114	

Les échantillons d'entreprises françaises et allemandes sont de taille similaire, alors que les entreprises britanniques sont plus nombreuses. La performance moyenne (mesurée par IB/(AT-IB)) des entreprises suit une tendance similaire dans les trois échantillons. Enfin, les écarts types année par année sont assez similaires. Ils sont plus faibles dans l'échantillon français.

Figure 110 : Caractéristiques des populations d'entreprises allemandes et britanniques

Évolution de la taille de l'échantillon



Évolution du résultat moyen (mis à l'échelle)	
Évolution de l'écart type	

Dans un premier temps, la distribution globale (centrée réduite année par année) est ajustée, puis dans un second temps, la partie droite et la partie gauche des distributions sont ajustées séparément.

a.2. Centrage des données

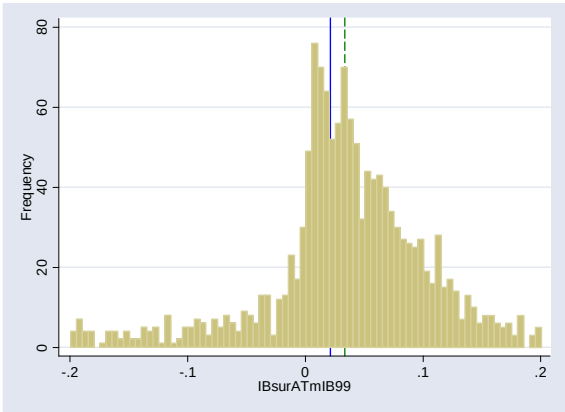
Avant de centrer et réduire, les deux indicateurs de tendance (médiane et moyenne) sont comparés. Pour les entreprises françaises, la médiane est plus proche du sommet de la distribution les années où médiane et moyenne sont sensiblement différentes. L'étude cherche à savoir si ce constat est valable pour les deux autres populations.

Tableau 77 : Comparaison moyenne / médiane des distributions annuelles ; population britannique

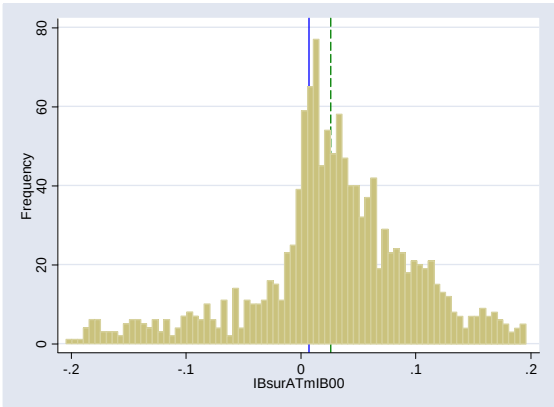
Entreprises britanniques	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
nombre d'observations	1112	1595	1664	1609	1539	1448	1398	1311	1220	1132
Moyenne	0.054	0.045	0.044	0.036	0.021	0.008	-0.019	-0.019	0.000	0.018
Médiane	0.052	0.047	0.048	0.044	0.033	0.027	0.017	0.020	0.023	0.026
Écart moy/med	-0.002	0.003	0.004	0.008	**0.012**	**0.019**	**0.035**	**0.039**	**0.023**	0.008

L'écart entre médiane et moyenne est visible (Tableau 77) en Grande Bretagne entre 1999 et 2003. Le centre des distributions est représenté graphiquement pour ces années-là, en traçant la moyenne en continu (bleu) et la médiane en pointillés (verts).

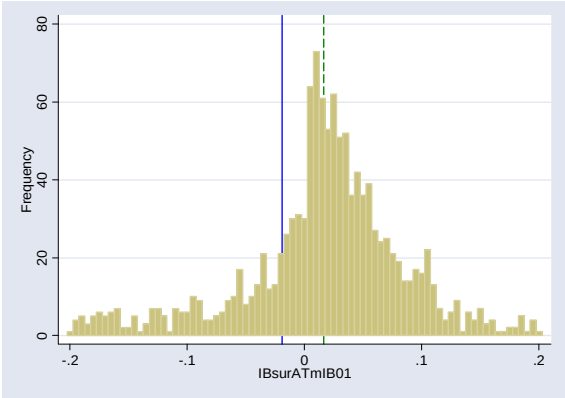
Figure 111 : Représentation graphique des distributions britanniques centrées-réduites ; 1999-2003



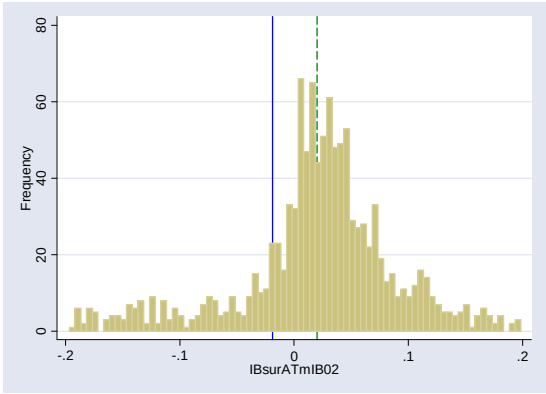
En 1999 la distribution est bimodale. La moyenne se confond avec l'irrégularité au zéro.



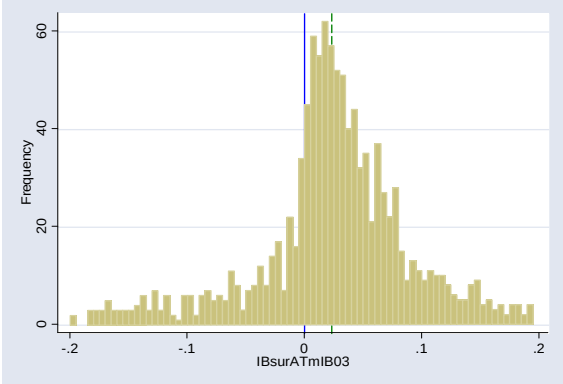
En 2000, la distribution est à nouveau bimodale. La moyenne se confond avec l'irrégularité au zéro.



En 2001, la médiane est manifestement située sur le sommet.



En 2002 à nouveau la médiane est plus proche du sommet que la moyenne.



En 2003 la médiane est plus proche du sommet que la moyenne

Pour mémoire, commande sous STATA :
histogram IBsurATmIB99 if IBsurATmIB99>-0.2 & IBsurATmIB99<0.2, width(0.005) frequency xline(0.021092, lcolor(blue)) xline(0.033292, lcolor(green) lpattern(dash))

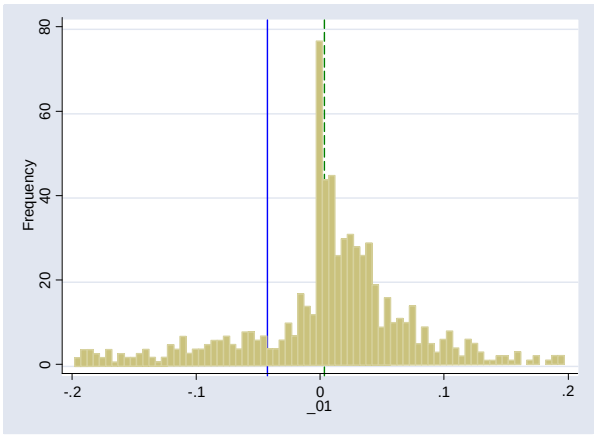
L'observation des distributions ne permet pas toujours de trancher de manière évidente. Par deux fois, en 1999 et 2000, les distributions sont bimodales. La moyenne et la médiane sont alors situées chacune sur un sommet. Mais ces années-là, la moyenne se confond avec le seuil du résultat nul, seuil étudié et dont les effectifs sont suspects. Les trois autres années, la médiane apparaît nettement comme étant plus proche du sommet. Par ailleurs, la médiane est moins volatile que la moyenne. Cette analyse confirme donc sur un échantillon britannique les observations déjà réalisées sur l'échantillon français : la médiane est plus proche du sommet que la moyenne.

Tableau 78 : Comparaison moyenne / médiane des distributions annuelles ; population allemande

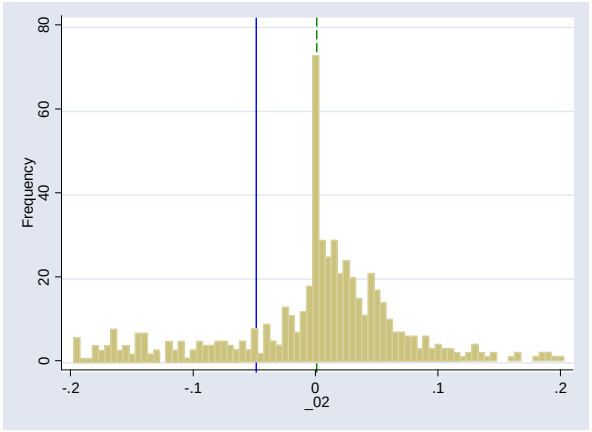
Entreprises allemandes	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**
nombre d'observations	365	526	598	737	886	899	837	721	679	631
Moyenne	0.026	0.028	0.035	0.033	0.017	0.006	-0.042	-0.049	-0.015	0.007
Médiane	0.018	0.018	0.025	0.025	0.016	0.015	0.003	0.001	0.004	0.014
Écart moy/med	-0.008	**-0.010**	**-0.010**	-0.008	-0.001	0.008	**0.046**	**0.049**	**0.019**	0.008

L'écart entre médiane et moyenne dans la population allemande est important essentiellement trois années : de 2001 à 2003. Le centre des distributions est représenté graphiquement pour ces années-là, en traçant la moyenne en continu (bleu) et la médiane en vert pointillés (verts).

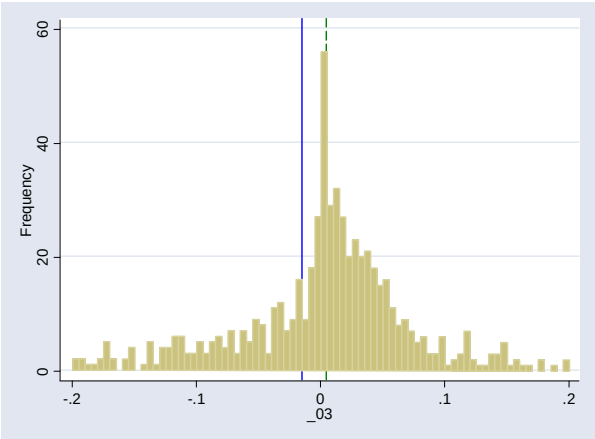
Figure 112 : Représentation graphique des distributions allemandes centrées-réduites ; 2001-2003



En 2001, la moyenne est loin du sommet.



En 2002, même constat.



En 2003, même constat.

Pour mémoire, commande sous STATA :
histogram _01 if _01>-0.2 & _01<0.2, width(0.005) frequency xline(-0.042479, lcolor(blue)) xline(0.003166, lcolor(green) lpattern(dash))

La médiane apparaît nettement comme étant plus proche du sommet. Ce résultat confirme et renforce donc les observations déjà réalisées sur les échantillons français et britannique. En conclusion, toutes les données sont centrées sur la médiane.

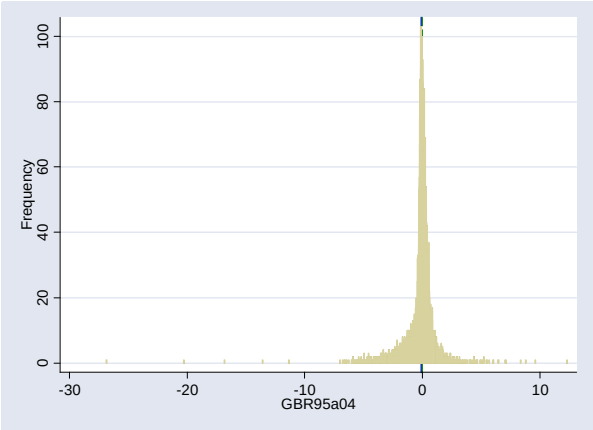
a.3. Description des résultats centrés réduits pour la Grande Bretagne

La seule observation graphique (Figure 113) des distributions globales laisse présumer une allure similaire aux distributions française et britannique, dissymétrique et fortement concentrée.

Tableau 79 : Paramètres de la distribution des résultats britanniques centrés, réduits et agrégés

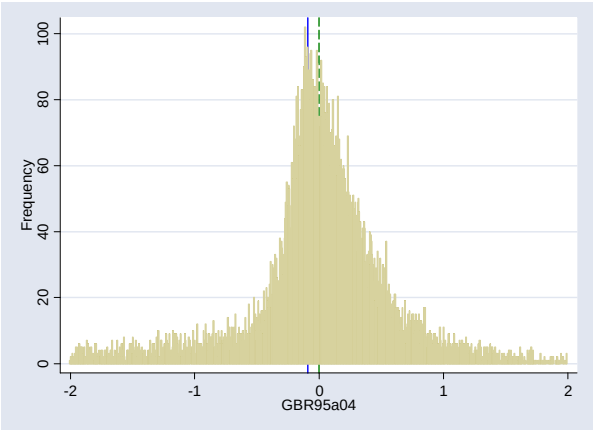
GBR95a04	N	mean	Médiane	Max	Min	Écart type	Skewness	kurtosis
	14028	-0.0895084	0	12.33841	-26.86409	1.00	-3.395523	69.71676

Figure 113 : Représentations graphiques de la distribution britannique centrée, réduite et agrégée

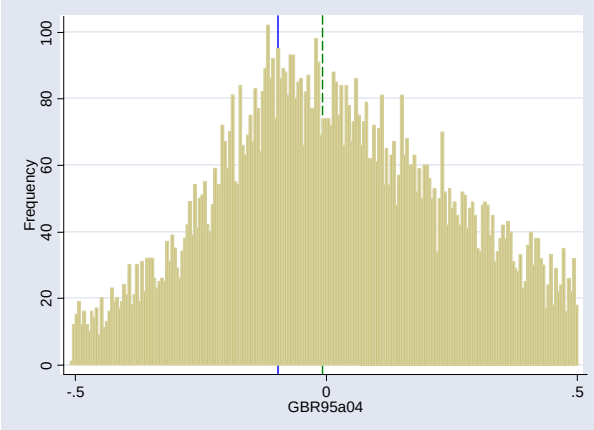


Distribution globale

Pour mémoire, sous Stata : histogram GBR95a04 if GBR95a04>-2 & GBR95a04<2, width(0.005) frequency xline(-0.0895084, lcolor(blue)) xline(0, lcolor(green) lpattern(dash)) (bin=799, start=-1.9994434, width=.005)



« Zoom » sur la partie centrale ; intervalle [-2 ;2]



« Zoom » sur la partie centrale ; intervalle [-0,5 ;0,5]

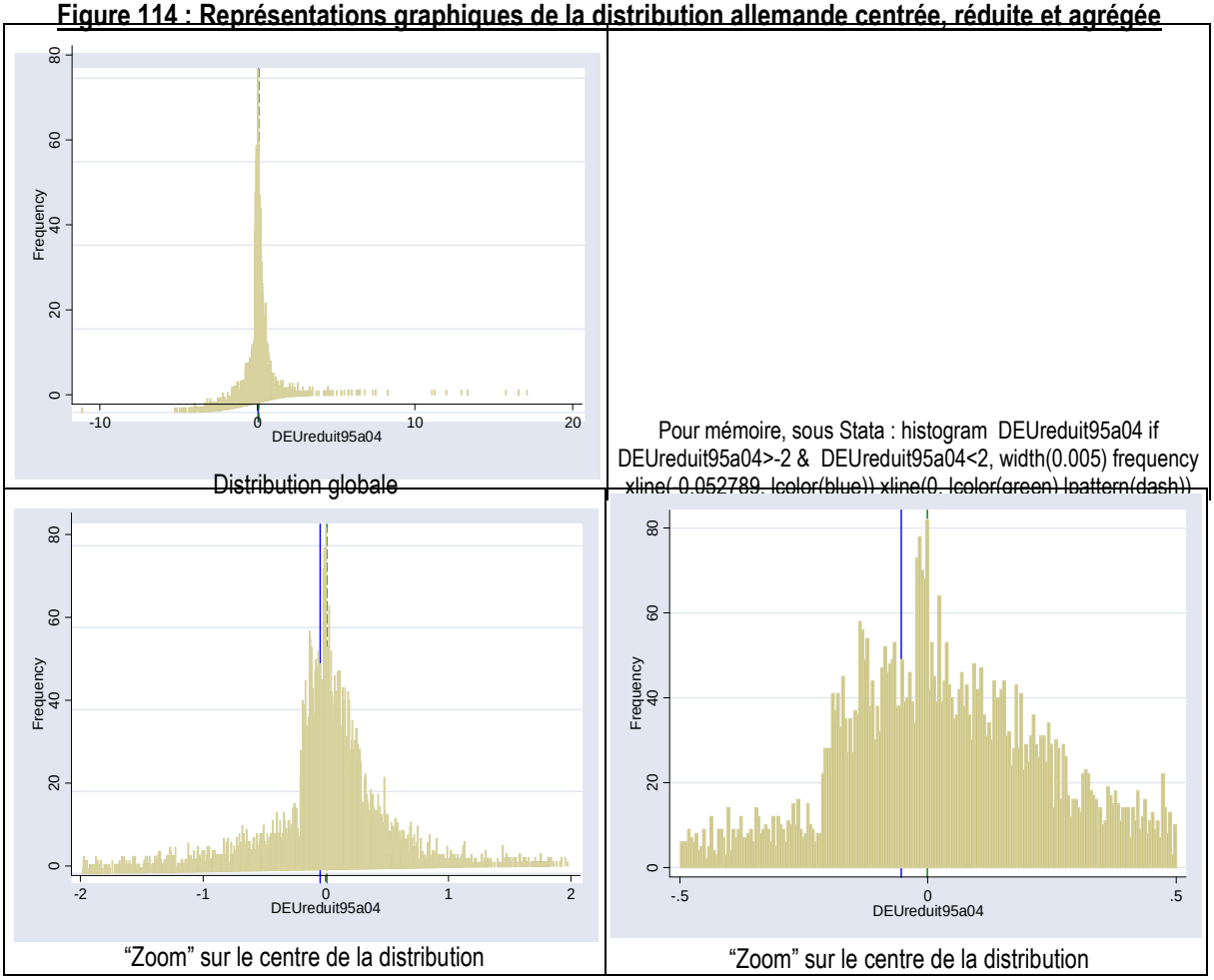
a.4. Description des résultats centrés réduits pour l'Allemagne

Il apparaît une particularité dans la distribution allemande : la dissymétrie est inversée. Ceci est dû à la présence de quelques valeurs extrêmes positives. L'allure de la partie centrale de la distribution est, elle, visuellement similaire aux distributions britanniques et françaises (Figure 114).

Tableau 80 : Paramètres de la distribution des résultats allemands centrés, réduits et agrégés

DEUreduit95a04	N	mean	Médiane	Max	Min	Écart type	Skewness	kurtosis
	6879	-0.052789	0	17.08891	-11.17674	1.00	3.088293	58.70341

Figure 114 : Représentations graphiques de la distribution allemande centrée, réduite et agrégée



Comme pour les entreprises françaises, deux ajustements sont réalisés pour chaque population. Dans un premier temps (sous-partie b), l'ajustement est réalisé pour la distribution globale (donc 2 ajustements, 1 pour chaque pays). Puis (sous-partie c) l'ajustement est réalisé sur les demi-distributions (donc 4 ajustements, 2 par pays). Au total, 6 ajustements sont réalisés.

b. Ajustement d'une loi de distribution globale

Les distributions des résultats des entreprises britanniques puis allemandes sont ajustées successivement.

b.1. Ajustement d'une loi de distribution des entreprises britanniques

Les ajustements sont présentés dans le Tableau 81.

Tableau 81 : Ajustement de la loi de distribution des résultats britanniques

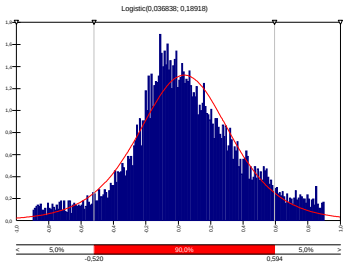
Loi	Distribution empirique	Logistique	Student	Normale	Weibull
Function		=RiskLogistic(-0,0081426; 0,40532)	=RiskStudent(6)	=RiskNormal(-0,089508; 1,0027)	=RiskWeibull(18,525; 27,499; RiskShift(-27,176))
Shift		N/A	N/A	N/A	-27,1762804177666
Parameter 1		-0,008142575696315	6	-0,08950836191189	18,5248392311466
Parameter 2		0,405315527401644	N/A	1,00270451396282	27,499477720811
Minimum	-26,864	-Infinity	-Infinity	-Infinity	-27,176
Maximum	12,338	+Infinity	+Infinity	+Infinity	+Infinity
Mean	-0,089508	-0,0081426	0,0000	-0,089508	-0,45811
Mode	-0,11206	-0,0081426	0,0000	-0,089508	0,24094
Median	0,0000	-0,0081426	0.0000	-0,089508	-0,21553
Std. Deviation	1,0027	0,73516	1,2247	1,0027	1,7833
Variance	1,0053	0,54046	1,5000	1,0054	3,1800
Skewness	-3,3955	0,0000	0,0000	0,0000	-0,8483
Kurtosis	69,7167	4,2000	6,0000	3,0000	4,1988
A-D Test Value		386,3	1038	+Infinity	+Infinity
P Value		< 0,005	N/A	< 0,005	N/A
Rank		**1**	2	3	4
K-S Test Value		0,1196	0,1806	0,2086	0,2865
P Value		< 0,01	N/A	< 0,01	N/A
Rank		**1**	2	3	4
Chi-Sq Test Value		5344	10988	10719	22939
P Value		0,0000	0,0000	0,0000	0,0000
Rank		**1**	3	2	4

Les scores aux tests ne sont pas très bons, mais ce qui est important, c'est que la loi la mieux classée est une loi Logistique, c'est-à-dire la même que celle retenue pour les entreprises françaises. La démarche qui consiste à trouver une loi "universelle" de distribution des résultats est donc confortée.

Comme cela a été réalisé dans le cas des entreprises françaises, l'ajustement a ensuite été réalisé sur la partie centrale de la distribution. Par tâtonnement, l'intervalle centré permettant le meilleur ajustement est cherché. Le meilleur ajustement (voir Tableau 82)

est obtenu sur les données comprises entre [-0,9 et 0,9]. La distance du test de K-S tombe alors à 0,02355 au lieu de 0,1196 (Tableau 81).

Tableau 82 : Meilleur ajustement de la partie centrale de la distribution des résultats britanniques (intervalle [-0,9 et 0,9])

Loi	Distribution empirique	Logistique	Normale
Function		=RiskLogistic(0,036838; 0,18918)	=RiskNormal(0,041517; 0,33754)
Illustration			
Shift		N/A	N/A
Parameter 1		3,68383143605538E-02	4,15174086091017E-02
Parameter 2		0,189180985628956	0,337539068651854
Minimum	-0,89952	-Infinity	-Infinity
Maximum	0,89995	+Infinity	+Infinity
Mean	0,041517	0,036838	0,041517
Mode	-0,063746	0,036838	0,041517
Median	0,020356	0,036838	0,041517
Std. Deviation	0,33754	0,34314	0,33754
Variance	0,11392	0,11774	0,11393
Skewness	0,0283	0,0000	0,0000
Kurtosis	3,1794	4,2000	3,0000
A-D Test Value		17,13	31,28
P Value		< 0,005	< 0,005
Rank		1	2
K-S Test Value		**0,02355**	0,03485
P Value		< 0,01	< 0,01
Rank		1	2
Chi-Sq Test Value		553,4	634,3
P Value		0,0000	0,0000
Rank		1	2

Tant que la zone étudiée (irrégularité au seuil du résultat nul) est inclue dans l'intervalle ajusté, supprimer les valeurs extrêmes n'est pas une contrainte. Il n'en demeure pas moins que la loi Logistique ne décrit pas convenablement la dissymétrie. En aucun cas elle ne peut être un bon ajusteur des queues de distribution. En revanche, ce qui doit être souligné, c'est que l'allure de la distribution des résultats des entreprises anglaises est ajustée au mieux par la même loi que celle ajustant la distribution des résultats des entreprises françaises.

b.2. Ajustement d'une loi de distribution des entreprises allemandes

Les ajustements fournis par le logiciel sont présentés dans le Tableau 83.

Tableau 83 : Ajustement de la distribution des résultats allemands

Loi	Distribution empirique	Logistique	Log-Logistique	Student	Inverse-Gaussienne	Normale	Gamma
Function		=RiskLogistic(-0,019314; 0,38721)	=RiskLogLogistic(-57,903; 57,883; 149,51)	=RiskStudent(6)	=RiskInvGauss(23,627; 13776,838; RiskShift(-23,680))	=RiskNormal(-0,052789; 1,0072)	=RiskGamma(335,93; 0,053540; RiskShift(-18,038478))
Shift		N/A	N/A	N/A	-23,67953369914	N/A	-18,0384782119
Parameter 1		-1,931386110E-02	-57,9027238241	6	23,62674471531	-0,05278898383	335,9285892483
Parameter 2		0,3872088195515	57,88252499482	N/A	13776,83824329	1,00724545287	5,35402159E-02
Parameter 3		N/A	149,5061543202	N/A	N/A	N/A	N/A
Minimum	-11,177	-Infinity	-57,903	-Infinity	-23,680	-Infinity	-18,038
Maximum	17,089	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	-0,052789	-0,019314	-0,015939	0,0000	-0,052789	-0,052789	-0,052789
Mode	-0,13634	-0,019314	-0,025378	0,0000	-0,11349	-0,052789	-0,10633
Median	0,0000	-0,019314	-0,020199	0.0000	-0,052789	-0,052789	-0,070633
Std. Deviation	1,0072	0,70232	0,70234	1,2247	0,97843	1,0072	0,98130
Variance	1,0144	0,49325	0,49328	1,5000	0,95733	1,0145	0,96296
Skewness	3,0883	0,0000	0,0583	0,0000	0,1242	0,0000	0,1091
Kurtosis	58,7034	4,2000	4,2084	6,0000	3,0257	3,0000	3,0179
A-D Test Value		+Infinity	+Infinity	590,2	+Infinity	+Infinity	+Infinity
P Value		< 0,005	N/A	N/A	N/A	< 0,005	N/A
Rank		4	5	1	3	6	2
K-S Test Value		0,1483	0,1486	0,2031	0,2053	0,2074	0,2127
P Value		< 0,01	N/A	N/A	N/A	< 0,01	N/A
Rank		**1**	**2**	3	4	5	6
Chi-Sq Test Value		3465	3510	7447	6899	7128	6982
P Value		0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Rank		**1**	**2**	6	3	5	4

On retrouve peu ou prou les mêmes lois. Aucune n'ajuste de manière réellement satisfaisante ; les ajustements des distributions dans leur globalité ne donnent pas de bons résultats. Cependant, la loi Logistique demeure toujours la meilleure en valeur relative. Cela conforte l'hypothèse que l'allure de la distribution des résultats publiés par les entreprises ne dépend pas du pays.

En réduisant progressivement la distribution à sa seule partie centrale, l'ajustement par la loi logistique est très sensiblement amélioré. Par tâtonnement, le meilleur ajustement est obtenu pour un intervalle [-0,6 ; 0,6]. La distance du test de K-S tombe à 0,02784 (Tableau 84) pour la loi logistique (au lieu de 0,1483 Tableau 83).

Tableau 84 : Ajustement de la partie centrale ([-0,6 ; 0,6]) de la distribution des résultats allemands

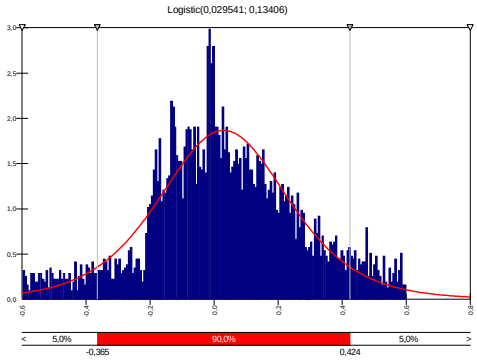
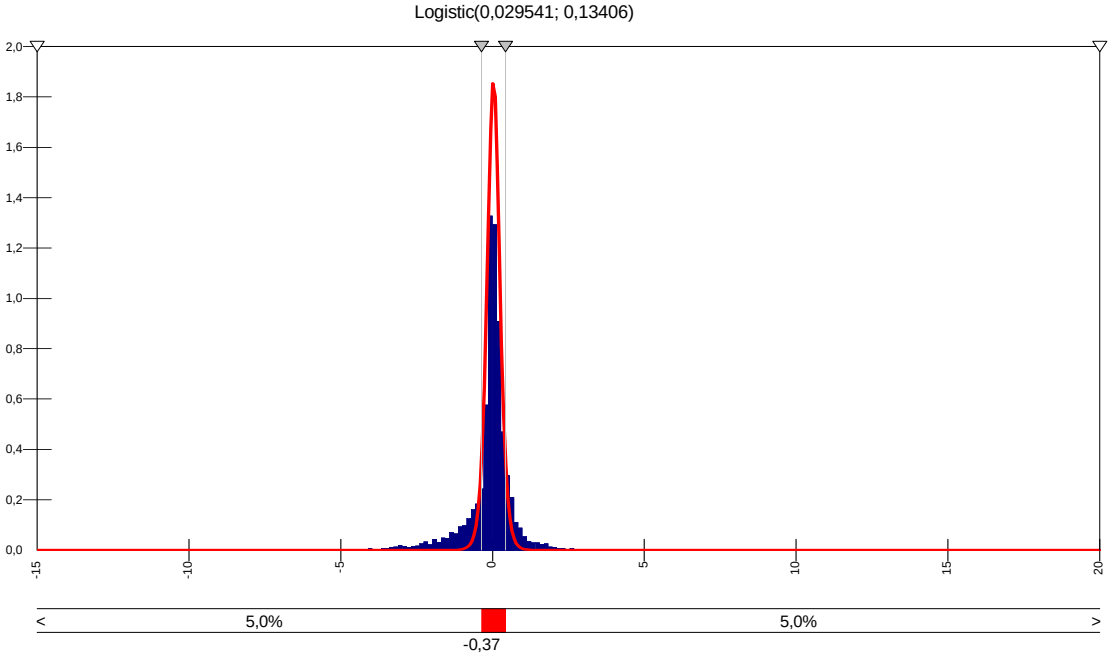
Loi	Logistique	Normale
Function	=RiskLogistic(0,029541; 0,13406)	=RiskNormal(0,030586; 0,23754)
Illustration		
Parameter 1	2,95405007636235E-02	3,05856206240487E-02
Parameter 2	0,134062994347601	0,237542903148811
K-S Test Value	0,02784	0,04039
P Value	< 0,01	< 0,01

Figure 115 : Comparaison de la loi ajustée sur la seule partie centrale des données, avec la distribution globale



La Figure 115 compare graphiquement la loi ajustée sur la seule partie centrale à la distribution dans son ensemble. Logiquement, la loi ajustée sur la seule partie centrale n'est pas adaptée pour ajuster la totalité de la distribution. Les queues de distribution sont ignorées.

b.3. Comparaison avec la distribution française

Le principal résultat de ce travail d'ajustement limité à la partie centrale des distributions est de constater que l'allure des distributions de résultats est similaire quelle que soit la nationalité des entreprises. Cela renforce l'intérêt de la démarche paramétrique de mesure des irrégularités de distribution. Il signifie en effet que la distribution des résultats publiés par les entreprises semble suivre une loi similaire quel que soit le pays étudié.

c. Ajustement d'une loi de distribution partielle (demi-distribution)

Comme pour l'échantillon d'entreprises françaises, les distributions de résultats britanniques et allemands sont scindés en deux demi-distributions : les résultats supérieurs à la médiane (valeurs positives après centrage), et les résultats inférieurs à la médiane (valeurs négatives transformées en positives en multipliant par -1). Chaque demi-distribution fait l'objet d'un ajustement. Les données strictement égales à zéro sont transformées en « 0,0000000001 » ou « -0,0000000001 » et intégrées dans chacun des deux échantillons.

c.1. Ajustement d'une loi de distribution partielle des entreprises britanniques

c.1.1. Ajustement des données positives

Les ajustements proposés par le logiciel pour la demi-distribution droite des résultats des entreprises britanniques sont présentés dans le Tableau 85.

Tableau 85 : Ajustement de la loi de distribution des résultats britanniques ; demi-distribution droite

	Distribution empirique	Loi monotones		Loi non monotones			
		Pareto2	Exponentielle	Log-Logistique	Log-Normale	Inverse-Gaussienne	Pearson 5
Function		=RiskPareto2(2,0960; 5,7441; RiskShift(0.0000))	=RiskExpon(0,44466; RiskShift(-6,3379E-05))	=RiskLogLogistic(-0,0065263; 0,26554; 1,5236)	=RiskLognorm2(-1,2849; 1,0472; RiskShift(-0,020049))	=RiskInvGauss(0,48902; 0,37112; RiskShift(-0,044355))	=RiskPearson5(2,0338; 0,60568; RiskShift(-0,097001))
Shift		0,0000000001	-0,0000633785682	N/A	-0,0200489783081	-0,0443552153838	-0,097001158567
Parameter 1		2,0959519161892	0,44466473635958	-0,006526265202	-1,2848588746733	0,48901995184338	2,0338156861881
Parameter 2		5,7440809215027	N/A	0,265544224268	1,04717545612313	0,37111894937180	0,6056810797733
Parameter 3		N/A	N/A	1,523626936995	N/A	N/A	N/A
Parameter 4		N/A	N/A	N/A	N/A	N/A	N/A
Parameter 5		N/A	N/A	N/A	N/A	N/A	N/A
Minimum	0.0000	0.0000	0.0000	-0,0065263	-0,020049	-0,044355	-0,097001
Maximum	12,338	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	0,44466	0,44180	0,44460	0,61439	0,45870	0,44466	0,48887
Mode	0.0000	0.0000	0.0000	0,088067	0,072369	0,072310	0,10264
Median	0,27622	0,26881	0,30815	0,25902	0,25664	0,25600	0,25680
Std. Deviation	0,59895	0,54723	0,44466	N/A	0,67603	0,56135	3,1860
Variance	0,35869	0,29946	0,19773	N/A	0,45701	0,31511	10,150
Skewness	5,5646	3,9684	2,0000	N/A	7,0517	3,4437	N/A
Kurtosis	62,1071	43,6072	9,0000	N/A	157,9045	22,7654	N/A
A-D Test Value		+Infinity	28,05	19,65	10,45	10,37	17,25
P Value		N/A	< 0,01	N/A	N/A	N/A	N/A
Rank		6	5	4	2	1	3
K-S Test Value		0,02580	0,04737	0,02617	0,02843	0,03018	0,03160
P Value		N/A	< 0,01	N/A	N/A	N/A	N/A
Rank		**1**	6	**2**	**3**	4	5
Chi-Sq Test Value		114,0	204,8	318,7	178,8	186,5	445,1
P Value		8,8814E-05	0,0000	0,0000	0,0000	0,0000	0,0000
Rank		**1**	4	5	**2**	**3**	6

On retrouve pour la population britannique les mêmes résultats que ceux trouvés sur la population française. C'est en soi un résultat intéressant. Les performances des ajustements sont nettement meilleures que pour l'ajustement global. Les lois bien classées sont les mêmes, confirmant que l'allure de la distribution des résultats est similaire quel que soit le pays. La loi la mieux classée, loi de Pareto2, est une loi monotone.

c.1.2. Ajustement des données négatives

Les ajustements de la demi-distribution gauche des résultats des entreprises britanniques sont les suivants :

Tableau 86 : Ajustement de la loi de distribution des résultats britanniques ; demi-distribution gauche

	Distribution empirique	Lois non monotones			Lois monotones	
		Inverse-Gaussienne	Log-Normale	Pearson 5	Pareto2	Exponentielle
Function		=RiskInvGauss(0,64764; 0,21589; RiskShift(-0,024013))	=RiskLognorm2(-1,3497; 1,3750; RiskShift(-0,0042527))	=RiskPearson5(1,1256; 0,22711; RiskShift(-0,043301))	=RiskPareto2(0,55406; 1,7796; RiskShift(0.0000))	=RiskExpon(0,62363; RiskShift(-8,8887E-05))
Shift		-2,401273127718E-02	-4,252709247119E-03	-4,330066244742E-02	0,0000000001	-8,888679127569E-05
Parameter 1		0,647643160567431	-1,34967040716479	1,12561676954725	0,554059617019316	0,62363042919025
Parameter 2		0,215893697625119	1,37502105861265	0,227109251321946	1,77959172373032	N/A
Minimum	0.0000	-0,024013	-0,0042527	-0,043301	0.0000	0.0000
Maximum	26,864	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	0,62363	0,62363	0,66317	1,7647	0,71070	0,62354
Mode	0.0000	0,047085	0,034897	0,063543	0.0000	0.0000
Median	0,23362	0,24526	0,25507	0,23528	0,26386	0,43218
Std. Deviation	1,0397	1,1217	1,5828	N/A	N/A	0,62363
Variance	1,0808	1,2583	2,5052	N/A	N/A	0,38891
Skewness	6,1249	5,1960	20,4513	N/A	N/A	2,0000
Kurtosis	96,7463	47,9974	2634,9835	N/A	N/A	9,0000
A-D Test Value		9,853	14,66	15,47	+Infinity	+Infinity
P Value		N/A	N/A	N/A	N/A	< 0,01
Rank		**1**	**2**	**3**	5	4
K-S Test Value		0,02615	0,03540	0,04055	0,04494	0,2002
P Value		N/A	N/A	N/A	N/A	< 0,01
Rank		**1**	**2**	**3**	4	5
Chi-Sq Test Value		296,0	381,7	501,2	524,3	2260
P Value		0,0000	0,0000	0,0000	0,0000	0,0000
Rank		**1**	**2**	**3**	4	5

Ce sont les mêmes lois qui sont proposées, mais le classement est légèrement différent. La loi de Pareto2 apparaît comme moins bien classée que l'inverse Gaussienne, mais la distance du test de K-S est du même ordre de grandeur que la loi la mieux classée (0,045 contre 0,026). La loi Pareto2 est la loi monotone la mieux classée. Elle est sensiblement meilleure que la loi exponentielle (0,045 contre 0,200 pour l'Inverse Gaussienne). L'étude de la population britannique conduit donc aux mêmes conclusions que l'étude de la population française.

c.2. Ajustement d'une loi de distribution partielle des entreprises allemandes

c.2.1. Ajustement des données positives

Les ajustements de la demi-distribution droite des résultats des entreprises allemandes sont les suivants :

Tableau 87 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution droite

	Distribution empirique	Lois non monotones				Lois monotones	
		Log-Logistique	Pearson5	LogNormale	Inverse-Gaussienne	Pareto2	Exponentielle
Function		=RiskLogLogistic(-0,0030427; 0,21630; 1,3950)	=RiskPearson5(1,6037; 0,34927; RiskShift(-0,062734))	=RiskLognorm2(-1,5006; 1,1877; RiskShift(-0,010496))	=RiskInvGauss(0,47788; 0,22999; RiskShift(-0,028398))	=RiskPareto2(0,77410; 2,7842; RiskShift(0.0000))	=RiskExpon(0,44948; RiskShift(-0,00013055))
Shift		N/A	0,062734178508109	-0,01049618183	-0,028397831684195	0,0000000001	-0,0001305494201
Parameter 1		-0,003,042726895	1,60373184449743	-1,50058607459	0,477879829418932	0,77410034889761	0,4494819976347
Parameter 2		0,2162984108615	0,349270983325555	1,187726990751	0,229987334385414	2,78420734600835	N/A
Parameter 3		1,395014648891	N/A	N/A	N/A	N/A	N/A
Minimum	0.0000	-0,0030427	-0,062734	-0,010496	-0,028398	0.0000	-0,00013055
Maximum	17,089	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	0,44948	0,62402	0,51579	0,44098	0,44948	0,43386	0,44935
Mode	0.0000	0,056385	0,071408	0,043910	0,046387	0.0000	-0,00013055
Median	0,22197	0,21326	0,20897	0,21250	0,21288	0,21883	0,31143
Std. Deviation	0,93540	N/A	N/A	0,79475	0,68885	0,81750	0,44948
Variance	0,87472	N/A	N/A	0,63162	0,47452	0,66830	0,20203
Skewness	9,0840	N/A	N/A	10,7360	4,3244	N/A	2,0000
Kurtosis	121,9179	N/A	N/A	467,3688	34,1678	N/A	9,0000
A-D Test Value	N/A	5,117	5,196	4,455	8,587	+Infinity	+Infinity
P Value	N/A	N/A	N/A	N/A	N/A	N/A	< 0,01
Rank	N/A	**2**	**3**	**1**	4	6	5
K-S Test Value	N/A	0,02156	0,02507	0,03016	0,03493	0,03539	0,1265
P Value	N/A	N/A	N/A	N/A	N/A	N/A	< 0,01
Rank	N/A	**1**	**2**	3	4	5	6
Chi-Sq Test Value	N/A	116,7	111,1	88,27	131,3	94,01	378,7
P Value	N/A	7,5237E-08	4,1476E-07	2,5287E-04	6,6305E-10	5,6032E-05	0,0000
Rank	N/A	4	**3**	**1**	5	**2**	6

On retrouve les mêmes lois que pour les distributions françaises et anglaises. La loi de Pareto2 est classée en deuxième position sur le critère du test du Chi2, mais elle est la loi monotone la mieux ajustée.

La distribution allemande se caractérise par la présence de valeurs extrêmes positives plus nombreuses que les populations françaises et britanniques. Pour contrôler les résultats, l'étude est reproduite en supprimant les valeurs extrêmes[171] :

Tableau 88 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution droite après suppression des valeurs extrêmes

	Distribution empirique	Lois monotones		Lois non monotones			
		Pareto2	Exponentielle	LogNormale	Inverse-Gaussienne	Log-Logistique	Pearson5
Function	N/A	=RiskPareto2(2,4968; 8,2350; RiskShift(0.0000))	=RiskExpon(0,34468; RiskShift(-0,00010243))	=RiskLognorm2(-1,5037; 1,0364; RiskShift(-0,018198))	=RiskInvGauss(0,38386; 0,31495; RiskShift(-0,039185))	=RiskLogLogistic(-0,0061773; 0,21148; 1,5171)	=RiskPearson5(2,0571; 0,49318; RiskShift(-0,080083))
Shift	N/A	0,0000000001	-0,0001024307248	-0,018198460334	-0,0391847921938	N/A	-0,080083297902
Parameter 1	N/A	2,4967691042046	0,3446797253682	-1,5036640517856	0,3838645176620	-0,006177265678	2,057050948370
Parameter 2	N/A	8,23495194839166	N/A	1,036370035827	0,31495268519551	0,2114824429383	0,4931794900317
Parameter 3	N/A	N/A	N/A	N/A	N/A	1,517068927585	N/A
Minimum	0.0000	0.0000	-0,00010243	-0,018198	-0,039185	-0,0061773	-0,080083
Maximum	2,3125	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	0,34468	0,34510	0,34458	0,36216	0,34468	0,49287	0,38648
Mode	0.0000	0.0000	-0,00010243	0,057748	0,058939	0,068330	0,081242
Median	0,21519	0,21925	0,23881	0,20412	0,20346	0,20531	0,20417
Std. Deviation	0,38566	0,39660	0,34468	0,52804	0,42378	N/A	1,9533
Variance	0,14869	0,15729	0,11880	0,27882	0,17959	N/A	3,8155
Skewness	2,2161	3,0700	2,0000	6,8403	3,3120	N/A	N/A
Kurtosis	8,5972	21,8924	9,0000	146,2963	21,2820	N/A	N/A
A-D Test Value	N/A	+Infinity	10,01	4,470	4,156	8,839	7,743
P Value	N/A	N/A	< 0,01	N/A	N/A	N/A	N/A
Rank	N/A	6	5	**2**	**1**	4	3
K-S Test Value	N/A	0,02070	0,04755	0,02362	0,02528	0,02804	0,02928
P Value	N/A	N/A	< 0,01	N/A	N/A	N/A	N/A
Rank	N/A	**1**	6	**2**	**3**	4	5
Chi-Sq Test Value	N/A	67,42	133,8	93,21	83,25	176,6	176,3
P Value	N/A	0,0270	2,9171E-10	6,9486E-05	8,7934E-04	0,0000	0,0000
Rank	N/A	**1**	4	**3**	**2**	6	5

La loi Pareto2 devient alors le meilleur ajusteur, et le résultat déjà obtenu pour l'échantillon français et britannique est confirmé. Son résultat au test de K-S est légèrement meilleur que la meilleure des lois non monotones ajustée sans supprimer les valeurs extrêmes.

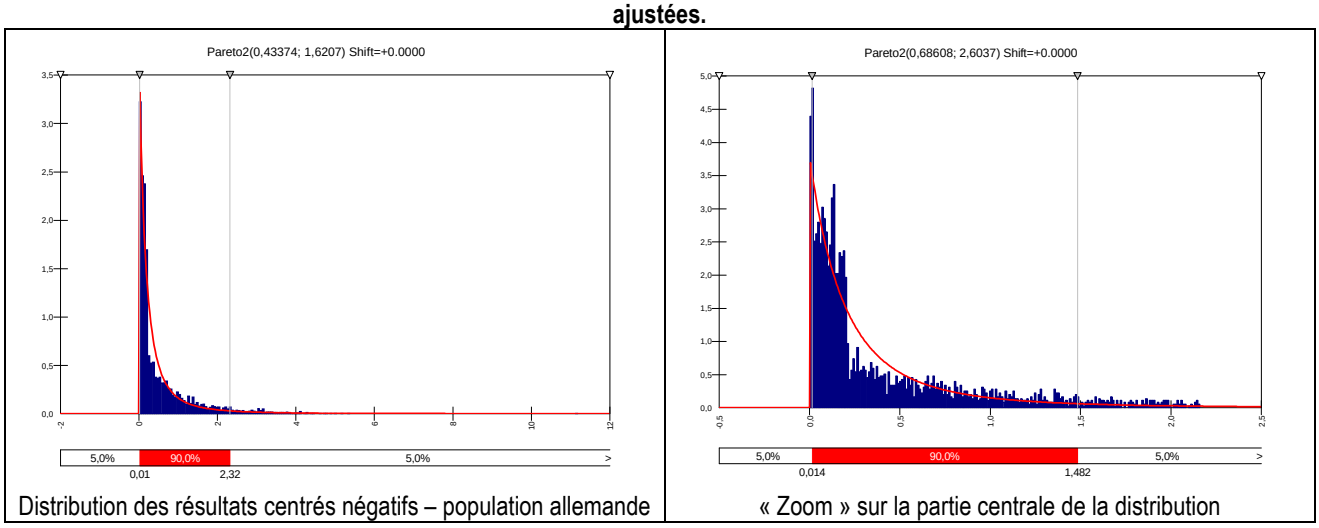
[171] Option « filtrer les données au-delà de deux écarts types de la moyenne » dans le logiciel.

c.2.2. Ajustement des données négatives

Visuellement (sur la Figure 116), il apparaît que l'ajustement des valeurs négatives pour les entreprises allemandes est plus difficile à trouver. La « pointe » de la distribution est très épaisse.

Figure 116 : Représentation graphique de la demi-distribution gauche des sociétés allemandes

Les données ont été préalablement multipliées par (-1) pour les transformer en valeurs positives afin d'être ajustées.



Distribution des résultats centrés négatifs – population allemande | « Zoom » sur la partie centrale de la distribution

Les ajustements de la demi-distribution gauche des résultats des entreprises allemandes sont présentés dans le Tableau 89 :

Tableau 89 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution gauche

	Distribution empirique	Lois non monotones			Lois monotones	
		Log-Normale	Inverse-Gaussienne	Pearson5	Pareto2	Exponentielle
Function		=RiskLognorm2(-1,5291; 1,5066; RiskShift(-0,0033292))	=RiskInvGauss(0,57872; 0,17345; RiskShift(-0,023770))	=RiskPearson5(1,0088; 0,16681; RiskShift(-0,038593))	=RiskPareto2(0,43374; 1,6207; RiskShift(0.0000))	=RiskExpon(0,55495; RiskShift(-0,00016118))
Shift		-0,003329213944167	-0,0237695924029314	-0,038592632639274	0,0000000001	-0,0001611827752421
Parameter 1		-1,52912330908968	0,57872223196166	1,00882661236946	0,433736551197454	0,554952639458728
Parameter 2		1,50663222828468	0,17344509744655	0,166809652513484	1,62071110760898	N/A
Minimum	0.0000	-0,0033292	-0,023770	-0,038593	0.0000	-0,00016118
Maximum	11,177	+Infinity	+Infinity	+Infinity	+Infinity	+Infinity
Mean	0,55495	0,67092	0,55495	18,860	0,69877	0,55479
Mode	0,13634	0,019063	0,033480	0,044446	0.0000	-0,00016118
Median	0,18895	0,21340	0,20231	0,19913	0,23148	0,38450
Std. Deviation	0,80483	1,9863	1,0571	N/A	N/A	0,55495
Variance	0,64757	3,9455	1,1175	N/A	N/A	0,30797
Skewness	2,8212	34,4056	5,4799	N/A	N/A	2,0000
Kurtosis	16,9073	10867,3410	53,0495	N/A	N/A	9,0000
A-D Test Value		**10,30**	**14,23**	**19,56**	**+Infinity**	**262,3**
P Value		N/A	N/A	N/A	N/A	< 0,01
Rank		1	2	3	5	4
K-S Test Value		**0,04281**	**0,05204**	**0,06008**	**0,06638**	**0,2233**
P Value		N/A	N/A	N/A	N/A	< 0,01
Rank		1	2	3	4	5
Chi-Sq Test Value		**389,4**	**383,9**	**552,9**	**419,6**	**1342**
P Value		0,0000	0,0000	0,0000	0,0000	0,0000
Rank		2	1	4	3	5

Les ajustements sont globalement moins bons. Mais le logiciel propose les mêmes lois. Donc malgré un aspect atypique, l'allure générale de la distribution reste similaire. La loi Pareto2 reste le meilleur ajusteur des lois monotones.

c.3. Comparaison avec la distribution française

En conclusion, le principal résultat de l'étude est que l'allure de la distribution des résultats est similaire quel que soit le pays étudié. La loi monotone qui ajuste le mieux les demi-distributions centrées sur la médiane est la loi Pareto2, qui est plus « pointue » que la loi exponentielle.

5. Conclusion et apports de la démarche d'ajustement

Le travail mené dans le chapitre VIII permet de proposer une démarche paramétrique d'estimation des irrégularités de distribution des résultats annuels publiés par les entreprises. Il conclut que la distribution théorique des résultats inférieurs à la médiane (demi-distribution gauche) suit une loi Pareto2.

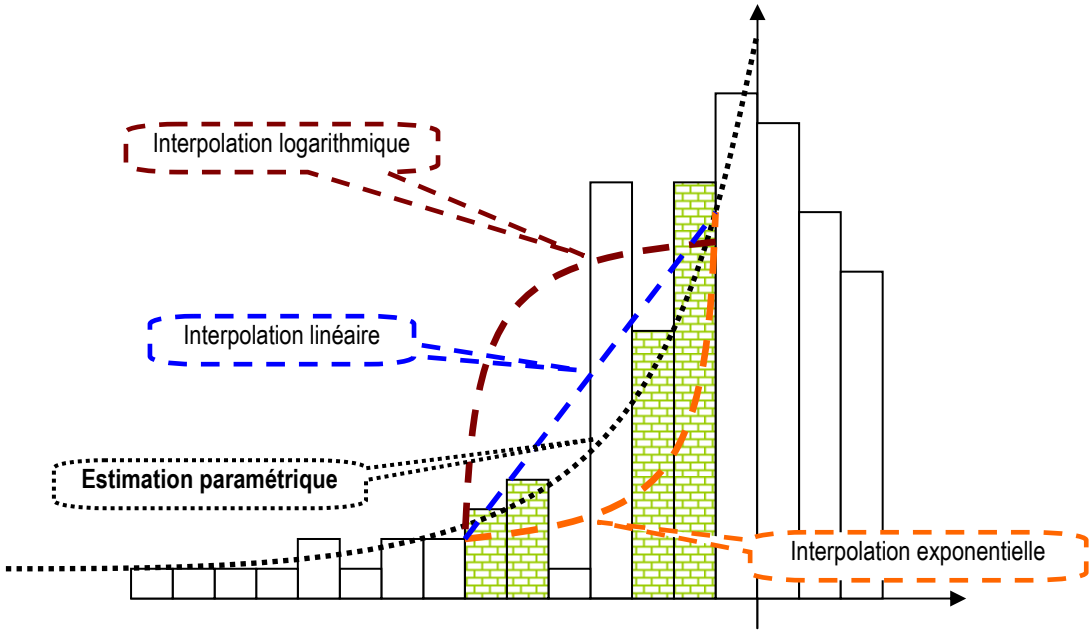
Une estimation moins précise peut être effectuée sur le centre de la distribution en retenant la loi Logistique (plus pointue que la loi Normale, mais moins pointue que la loi de Student). Ce résultat annexe mérite d'être souligné dans la mesure où l'étude de la partie centrale de la distribution peut sembler plus aisé dans le cadre de futures recherches. L'ajustement par la loi Pareto2 nécessite en effet le découpage de la distribution par rapport à la médiane, puis la transposition des données négatives (multiplication par « -1 »). Les variables extrêmes enfin peuvent influencer les calculs. Ce travail est lourd, et les graphes doivent être inversés avant d'être interprétés (puisque les données sont transposées). Tant que la zone étudiée (irrégularité au seuil du résultat nul) est inclue dans l'intervalle ajusté, supprimer les valeurs extrêmes n'est pas une contrainte. Mais une telle démarche interdit d'élargir l'intervalle d'étude au-delà des bornes fixées, ce qui peut être contraignant, notamment dans le cadre de l'étude des montants manipulés (comme cela sera vu dans le chapitre IX).

Les résultats apparaissent d'autant plus robustes qu'ils sont valables aussi bien pour les entreprises françaises que britanniques et allemandes. Ce constat renforce la validité de la démarche. Ce sont ces résultats qui sont appliqués dans l'étude menée dans le chapitre IX sur l'évolution des irrégularités des résultats des entreprises françaises de 1993 à 2004.

Enfin, l'intérêt de ces résultats mérite d'être mis en perspective avec les travaux menés dans la partie 2 de la thèse. Il a été mis en évidence qu'aucun fondement théorique ne permet de trancher en faveur d'une des différentes méthodes de mesure non paramétrique (principalement interpolation ou extrapolation, linéaire, exponentielle ou logarithmique). En fonction de l'emplacement de l'irrégularité dans la distribution ou de la largeur des intervalles, des arguments peuvent être retenus en faveur ou défaveur de chacune des méthodes (voir Figure 117).

Figure 117 : Comparaison graphique de la méthode d'estimation Pareto2 par rapport aux estimations non paramétriques

Sans réflexion sur l'allure générale de la loi, il est impossible de trancher entre interpolation linéaire, logarithmique ou exponentielle. Pourtant ces différentes mesures fournissent des résultats parfois très différents.



La conclusion retenue dans la partie 2 est que, en absence de critère objectif de sélection, l'interpolation linéaire est une mesure prudente qui fournit des mesures intermédiaires par rapport aux autres méthodes. Cependant, cette conclusion est un moindre mal. De ce point de vue, le résultat apporté ici (ajustement global de la distribution par la loi Pareto2) peut être considéré comme un progrès dans l'estimation des irrégularités de distribution des résultats publiés.

CHAPITRE IX - APPLICATION DANS LE CADRE D'UNE ÉTUDE LONGITUDINALE

Pour mettre en application l'instrument de mesure proposé dans le chapitre VIII, l'évolution de 1992 à 2004 de l'irrégularité au seuil du résultat nul dans l'échantillon d'entreprises françaises est étudiée. La section A mesure les irrégularités annuelles sur la largeur 0,05 selon les deux estimations paramétriques proposées dans le chapitre VIII (Logistique et Pareto2). Les résultats sont comparés aux résultats issus des méthodes non paramétriques de mesures d'irrégularités utilisées jusqu'alors dans la littérature (et développées dans la partie 2).

Mais l'intérêt du nouvel instrument de mesure est de permettre d'approfondir l'analyse des pratiques de gestion du résultat pour éviter les seuils en estimant le nombre total d'entreprises manipulatrices, et en mesurant les montants manipulés. L'étude de ces deux nouvelles dimensions fait l'objet de la section B.

A. Étude longitudinale de l'évolution des seuils en France

Les irrégularités sont mesurées selon la démarche paramétrique dans la sous-partie 1, puis les résultats sont comparés dans la sous-partie 2 aux mesures non paramétriques.

1. Mesures des irrégularités

Le chapitre VIII propose deux ajustements paramétriques : Logistique sur la partie centrale de la distribution et Pareto2 sur les demi-distributions. Ces deux lois sont utilisées successivement (sous-partie a, puis b).

a. Mesures par la loi Logistique (hypothèse de symétrie)

a.1. La démarche générale

D'après les résultats du chapitre VIII, le meilleur ajustement de la partie centrale de la distribution centrée sur la médiane, soit l'intervalle [-0,8 ; 0,8] pour les entreprises françaises, est réalisé par une loi logistique. Pour estimer les effectifs attendus, et donc en déduire les irrégularités, seules les données comprises sur [m-0,8σ ; m+0,8σ] sont retenues chaque année, avec m = médiane annuelle et σ = écart type annuel.

Pour les comparer aux mesures réalisées par les démarches non paramétriques de la partie 2 de la thèse, les irrégularités à gauche et à droite du seuil résultat nul sont mesurées en retenant une largeur d'intervalle de 0,05.

a.2. Les mesures année par année

Les irrégularités sont mesurées année par année. Le détail des calculs est fourni en Annexe 12 : Mesures année par année avec une loi Logistique (hypothèse de symétrie). Seuls les calculs de l'année 1992 sont décrit ci-dessous à titre d'exemple pour illustrer la démarche mise en œuvre.

La première ligne du tableau ci-dessous récapitule les caractéristiques de la distribution, et en déduit les bornes d'étude (médiane +/- 0,8σ).

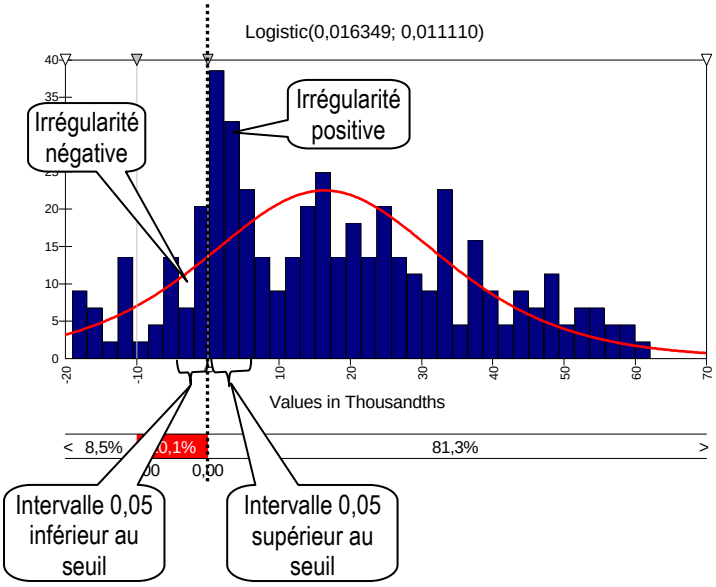
Tableau 90 : Ajustement de la distribution 1992 par une loi Logistique – partie 1 sur 3

				Détermination des bornes en 1992	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m – 0,8σ)	**borne sup** (m + 0,8σ)
291	0,028	0,0197	0,0545	-0,0239	0,0633

Les caractéristiques de la loi ajustée à la distribution annuelle sont ensuite précisées dans la partie gauche du tableau. Une représentation graphique illustre l'ajustement dans la partie droite du tableau.

Ajustement de la distribution 1992 par une loi Logistique – partie 2 sur 3

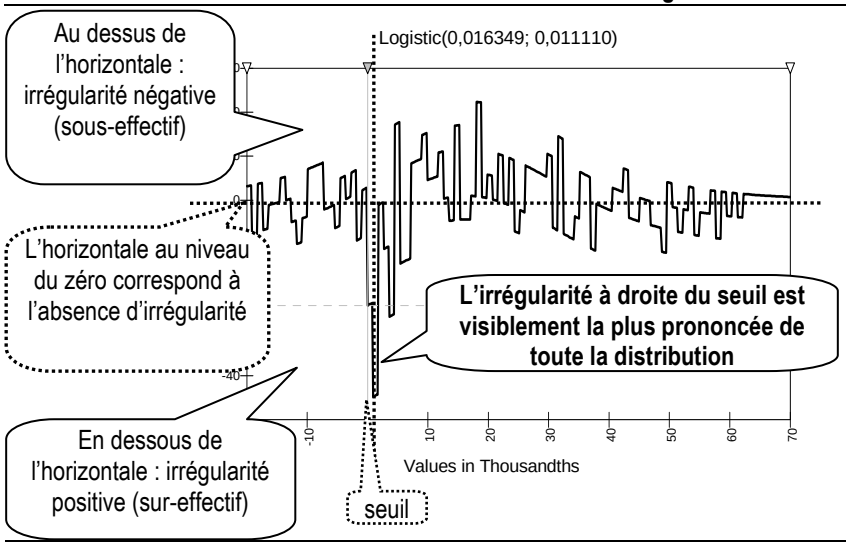
	Input	Logistic
Function		=RiskLogistic(0,016349; 0,011110)
Parameter 1		0,0163492313502607
Parameter 2		0,01110988660224
Minimum	-0,019100	-Infinity
Maximum	0,062100	+Infinity
Mean	0,017218	0,016349
Mode	0,00060000	0,016349
Median	0,015300	0,016349
Std. Deviation	0,018958	0,020151
Variance	0,00035765	0,00040607
Skewness	0,3231	0,0000
Kurtosis	2,3460	4,2000
A-D Test Value		1,686
P Value		< 0,005
Rank		7
K-S Test Value		0,08120
P Value		< 0,01
Rank		6
Chi-Sq Test Value		26,86
P Value		0,0201
Rank		6



Les effectifs théoriques (calculés à partir des paramètres de la loi décrite ci-dessus) et observés sont ensuite reportés dans la troisième partie du tableau. L'irrégularité en est déduite par soustraction, et est exprimée en pourcentage. Une représentation des écarts relatifs entre loi théorique et distribution empirique est présentée à titre d'illustration dans la partie gauche du tableau. Cette représentation graphique permet de comparer visuellement les irrégularités qui encadrent le seuil aux autres irrégularité présentes sur la distribution. Il est ainsi possible de constater que les irrégularités autour du seuil résultat nul sont en général les plus importantes de la distribution.

Ajustement de la distribution 1992 par une loi Logistique – partie 3 sur 3

Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	13	33
Ajusté	17.67	21.43
Irrégularité	-4,67	11,57
soit	-26.43%	**53.99%**

Ainsi, en 1992, l'irrégularité au dessus du résultat nul (54 %) apparaît comme étant la plus importante de toute les irrégularités positives de la distribution. L'irrégularité en dessous du résultat nul est peu prononcée mais négative (-26 %) conformément à la théorie.

Les ajustements ont été reproduits 13 fois (pour les 13 années), et les irrégularités à gauche et à droite du seuil résultat nul ont été mesurées. Une synthèse des résultats est présentée ci-dessous.

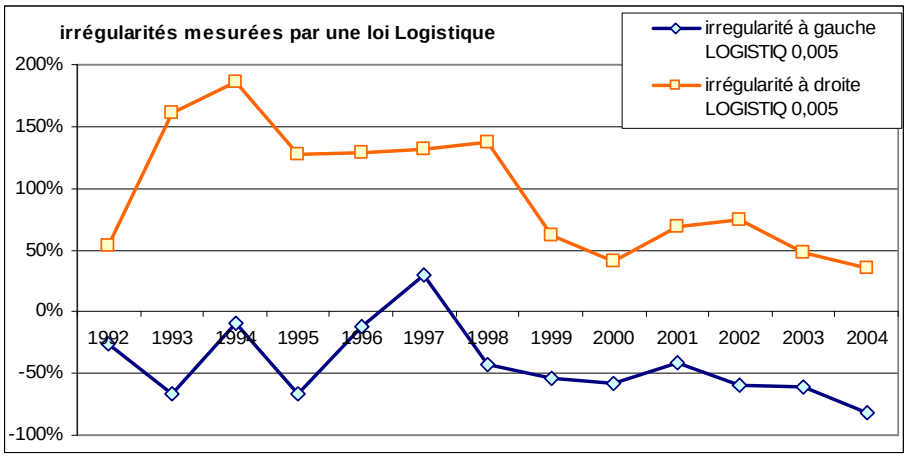
a.3. La synthèse des résultats

Le Tableau 91 récapitule les irrégularités mesurées sur les intervalles de largeur 0,005 encadrant le résultat nul, en ajustant la distribution par une loi Logistique. Ces mesures sont ensuite représentées graphiquement.

Tableau 91 : Irrégularités mesurées par l'ajustement Logistique

	effectif théorique gauche	effectif observé gauche	Irrégularité à gauche[172] [−0,05 ; 0]	effectif théorique à droite	effectif observé à droite	Irrégularité à droite [0 ; 0,05]
1992	17.67	13	-26.43%	21.43	33	53.99%
1993	11.97	4	-66.58%	14.57	38	160.81%
1994	11	10	-9.09%	14.34	41	185.91%
1995	11.97	4	-66.58%	17.55	40	127.92%
1996	15.97	14	-12.34%	22.74	52	128.67%
1997	16.16	21	29.95%	20.33	47	131.19%
1998	15.58	9	-42.23%	19.41	46	136.99%
1999	17.35	8	-53.89%	21.05	34	61.52%
2000	21.18	9	-57.51%	24.79	35	41.19%
2001	22.16	13	-41.34%	24.88	42	68.81%
2002	22.3	9	-59.64%	24.73	43	73.88%
2003	20.6	8	-61.17%	23.65	35	47.99%
2004	16.61	3	-81.94%	19.3	26	34.72%

Figure 118 : Représentation graphique des irrégularités mesurées par l'ajustement Logistique



Conformément à la théorie, les distributions annuelles présentent une irrégularité positive (surnombre d'entreprises faiblement bénéficiaires) au dessus du seuil du résultat nul, et une irrégularité négative (insuffisance du nombre d'entreprises faiblement déficitaires) au dessous (sauf pour une seule et unique année : 1997). L'ampleur de la première irrégularité tend à diminuer entre 1992 et 2004, alors que la seconde tend à devenir plus importante après 1998. Ces résultats mesurés en ajustant la partie centrale de la distribution par une loi Logistique méritent d'être confirmés en ajustant la demi-distribution par une loi Pareto2.

[172] Les irrégularités sont exprimées en valeur relative, c'est-à-dire irrégularité / effectif théorique.

b. Mesures par la loi Pareto2 (hypothèse de non symétrie)

b.1. La démarche générale

La première ligne du tableau ci-dessous récapitule les caractéristiques de la distribution, et en déduit les bornes d'étude (médiane +/- 0,8σ).

Après avoir mesuré les irrégularités en ajustant la partie centrale de la distribution par une loi Logistique, les mêmes irrégularités sont mesurées en ajustant la partie gauche de la distribution (transposée en multipliant par « -1 ») par une loi Pareto2.

b.2. Les mesures année par année

Les irrégularités sont une fois de plus mesurées année par année. Le détail des calculs est fourni en Annexe 13 : Mesures année par année avec une loi Pareto 2 (hypothèse de non symétrie). Seuls les calculs de l'année 1992 sont décrit ci-dessous à titre d'exemple pour illustrer la démarche mise en œuvre.

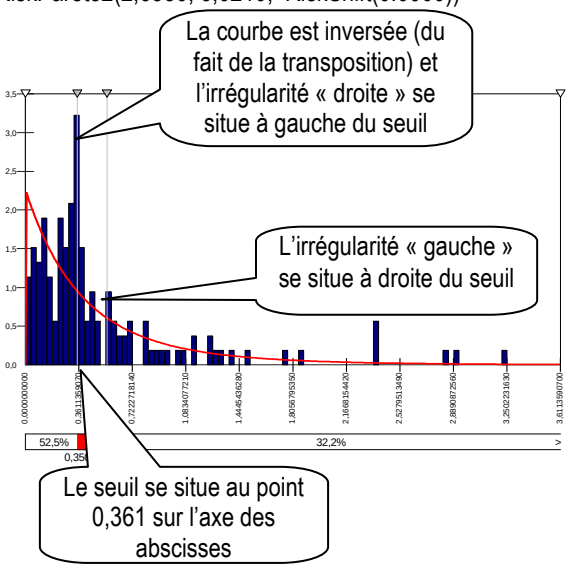
Année 1992

Pour être en mesure d'interpréter les représentations graphiques, il est important de rappeler que la partie gauche de la distribution est transformée : centrée sur la médiane et réduite par l'écart type, puis multipliée par (-1). La distribution à ajuster n'est donc composée que de valeurs positives. En 1992, le résultat nul dans la distribution initiale se retrouve, sur l'axe des abscisses, au point 0,361[173] dans la distribution transformées. Graphiquement, la demi-distribution transformée est représentée symétriquement par rapport à la distribution globale. L'irrégularité droite (sur la distribution globale) se situe donc à gauche du seuil (sur la demi-distribution transformée). De même, l'irrégularité gauche se situe à droite du seuil.

[173] [(résultat nul – médiane)/écart type] x (-1)= [(0-0,019673)/0,05447] x (-1) = 0,3611713

Tableau 92 : Ajustement de la demi-distribution gauche 1992 par une loi Pareto2

	Distribution de IB/(AT-IB) en 1992 (rappel)	Input = demi-distribution centrée, réduite, et transposée	Pareto2
N	291	146	
Function			=RiskPareto2(2,6586; 6,0210; RiskShift(0.0000))
Shift			1E-10
Parameter 1			2,658574192
Parameter 2			6,021048825
Minimum	-0,2224	0.0000	0.0000
Maximum	0,2136	4,4446	+Infinity
Mean	0,02806271	0,53112	0,52949
Mode		0,11116	0.0000
Median	0,019672663	0,33463	0,32437
Std. Deviation	0,054474404	0,66217	0,64792
Variance		0,43546	0,4198
Skewness		3,0576	3,7985
Kurtosis		14,1826	38,3198
A-D Test Value			+Infinity
P Value			N/A
Rank			1
K-S Test Value			0,1139
P Value			N/A
Rank			1
Chi-Sq Test Value			49,53
P Value			1,6864E-06
Rank			1



Position du résultat nul	**0,361135907**
[(résultat nul – médiane)/écart type] x (-1)	

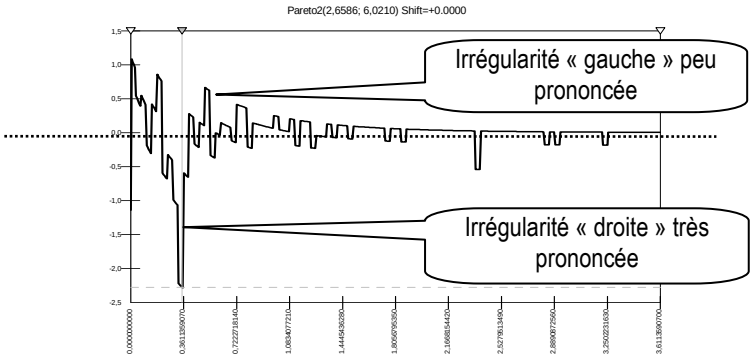
L'irrégularité au seuil du résultat nul est observable graphiquement (lecture au seuil 0,3611 dans la distribution transformée). En revanche, il apparaît aussi que le sommet de la distribution (donc la partie gauche du graphique représentant la demi-distribution) semble moins bien ajusté par la loi. De manière générale, il apparaît que la loi Pareto2 ajuste moins bien les premières années de l'échantillon.

Autour du seuil du résultat nul, l'irrégularité à droite (donc à gauche du seuil 0,361 sur la distribution ajustée) correspondant aux entreprises faiblement bénéficiaires est très prononcée. En revanche, l'irrégularité à gauche (donc à droite du seuil 0,361 sur la distribution ajustée) correspondant aux entreprises faiblement déficitaires est beaucoup moins prononcée. Il apparaît cependant visuellement que les irrégularités sont conformes à la théorie puisqu'il y a une surreprésentation des entreprises faiblement bénéficiaires et une sous représentation des entreprises faiblement déficitaires.

L'étude de la distribution des erreurs permet de confirmer l'analyse graphique et de construire le graphe des irrégularités (Tableau 93). Sur ce graphe, il apparaît que

l'irrégularité à droite du résultat nul (donc à gauche du trait vertical correspondant au seuil 0,361 de la distribution transformée) est la plus forte de toute la distribution (le « trou » est le plus « profond » sur le graphique). En revanche, l'irrégularité à gauche du résultat nul est peu prononcée. Mais elle apparaît tout de même comme l'une des plus importantes parmi les irrégularités négatives puisque le « pic » sur le graphe est l'un des plus « hauts ».

Tableau 93 : Distribution des irrégularités en 1992 ; ajustement Pareto2



Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,36114	0,26935
Max	0,45292	0,36114
Observé	14	32
Ajusté	11,19	13,85
Irrégularité	2,81	18,15
soit	**25,1%**	**131,0%**

En conclusion, l'étude de ce graphique est très intéressante. Dans les méthodes non paramétriques, les auteurs se contentaient de mesurer localement l'irrégularité, et estimaient par un test statistique sa significativité. La mesure paramétrique permet une représentation relative de l'irrégularité. Elle peut être comparée aux autres irrégularités qui se manifestent sur l'ensemble de la distribution. Autrement dit, ce n'est pas tant la significativité de l'irrégularité qu'il est intéressant de mesurer, que le fait que l'irrégularité au seuil du résultat nul est la plus prononcée de toutes les irrégularités observées sur la distribution.

La synthèse des résultats

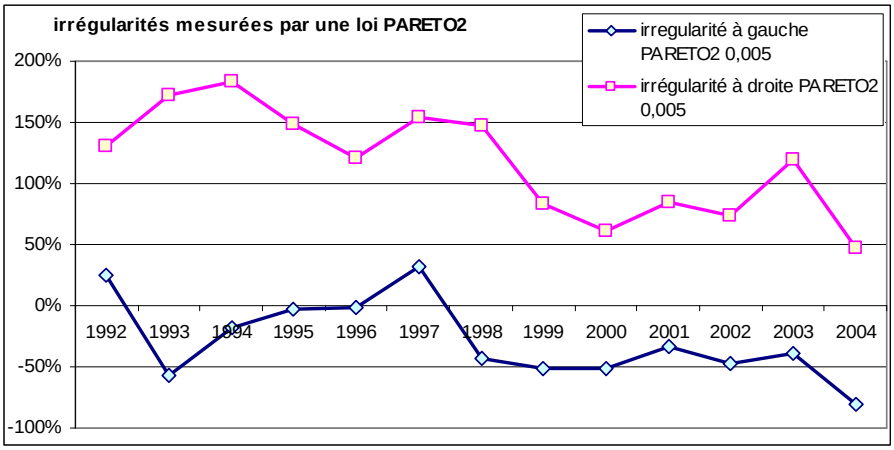
Les calculs en référence à une loi Pareto2 sur les intervalles de largeur 0,05 encadrant le résultat nul ont ensuite été reproduits sur les 13 années. Le Tableau 94 récapitule les irrégularités ainsi mesurées.

Tableau 94 : Irrégularités mesurées en ajustement la distribution par une loi Pareto2

PARETO2	effectif théorique gauche	effectif observé gauche	Irrégularité à gauche sur [−0,05 ; 0]
1992	11.19	14	0.25111707
1993	11.6	5	-0.56896552
1994	12.18	10	-0.17898194
1995	13.44	13	-0.0327381
1996	16.3	16	-0.01840491
1997	15.85	21	0.32492114
1998	15.98	9	-0.43679599
1999	16.48	8	-0.51456311
2000	18.39	9	-0.51060359
2001	19.43	13	-0.33093155
2002	20.68	11	-0.46808511
2003	14.63	9	-0.3848257
2004	15.11	3	-0.80145599

	effectif théorique à droite	effectif observé à droite	Irrégularité à droite sur [0 ;0,05]
1992	13.85	32	1.31046931
1993	13.96	38	1.72206304
1994	14.45	41	1.83737024
1995	16.46	41	1.490887
1996	19.43	43	1.21307257
1997	18.54	47	1.53505933
1998	18.63	46	1.4691358
1999	19.13	35	0.82958704
2000	21.7	35	0.61290323
2001	22.69	42	0.8510357
2002	24.12	42	0.74129353
2003	15.48	34	1.19638243
2004	17.65	26	0.47308782

Figure 119 : Représentation graphique des irrégularités par ajustement Pareto2



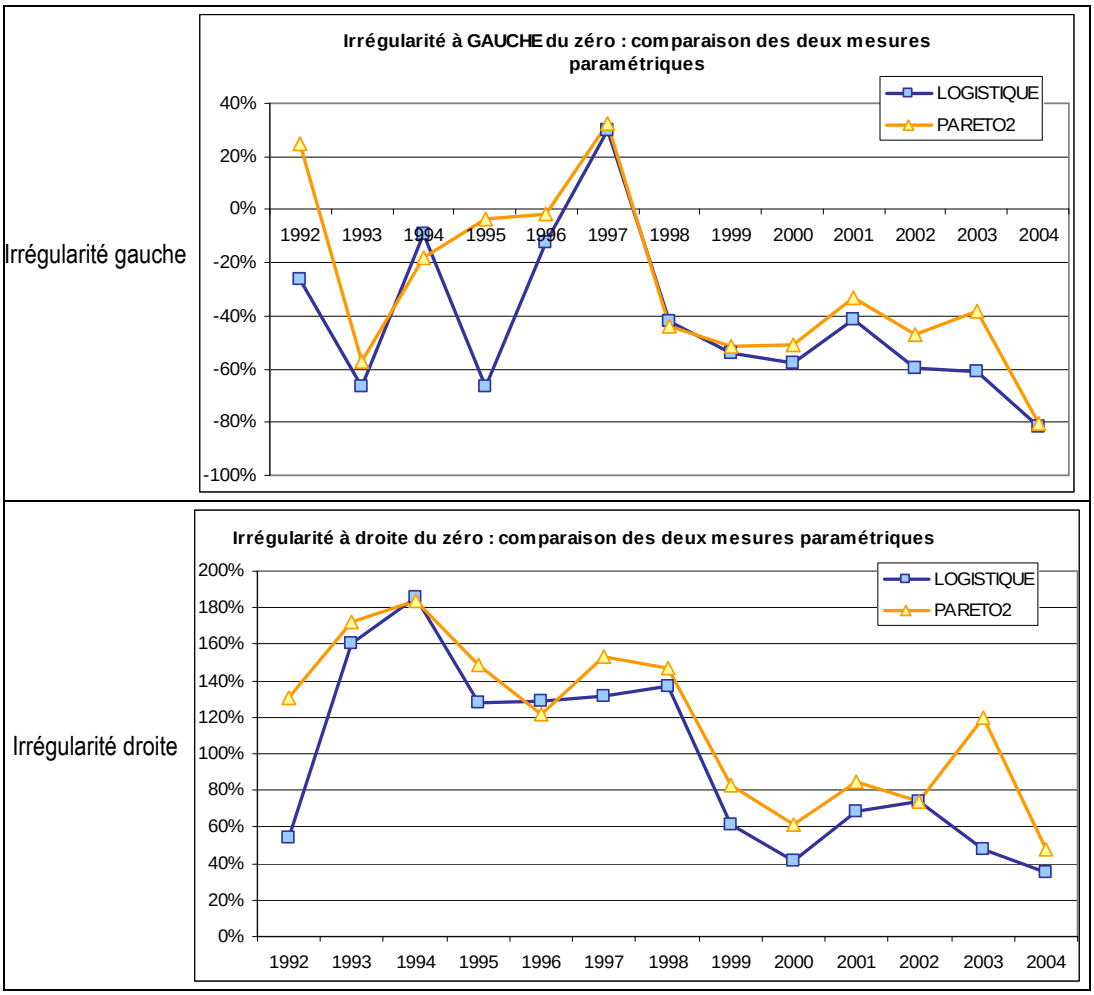
Les irrégularités montrent, conformément à la théorie, une irrégularité positive (surnombre d'entreprises faiblement bénéficiaires) au dessus du seuil du résultat nul, et une irrégularité négative (insuffisance du nombre d'entreprises faiblement déficitaires) au dessous. L'ampleur de la première irrégularité tend à diminuer entre 1992 et 2004, alors que la seconde, qui était peu importante voire inexistante entre 1992 et 1996, tend à devenir plus importante après 1998.

2. Comparaison des mesures d'irrégularité

Dans un premier temps, les mesures réalisées par les deux ajustements (Logistique et Pareto2) sont comparées. Puis ces mesures sont comparées avec les mesures issues des méthodes non paramétriques (mesurées dans la partie 2 de la thèse).

a. Comparaison des deux méthodes paramétriques

Figure 120 : Comparaison des irrégularités mesurées par ajustement Logistique et Pareto2



Il est intéressant de souligner la forte corrélation des mesures effectuées selon les deux méthodes[174]. Les deux démarches paramétriques paraissent donc très cohérentes. Elles font toutes deux apparaître une irrégularité négative à gauche du seuil, et une irrégularité

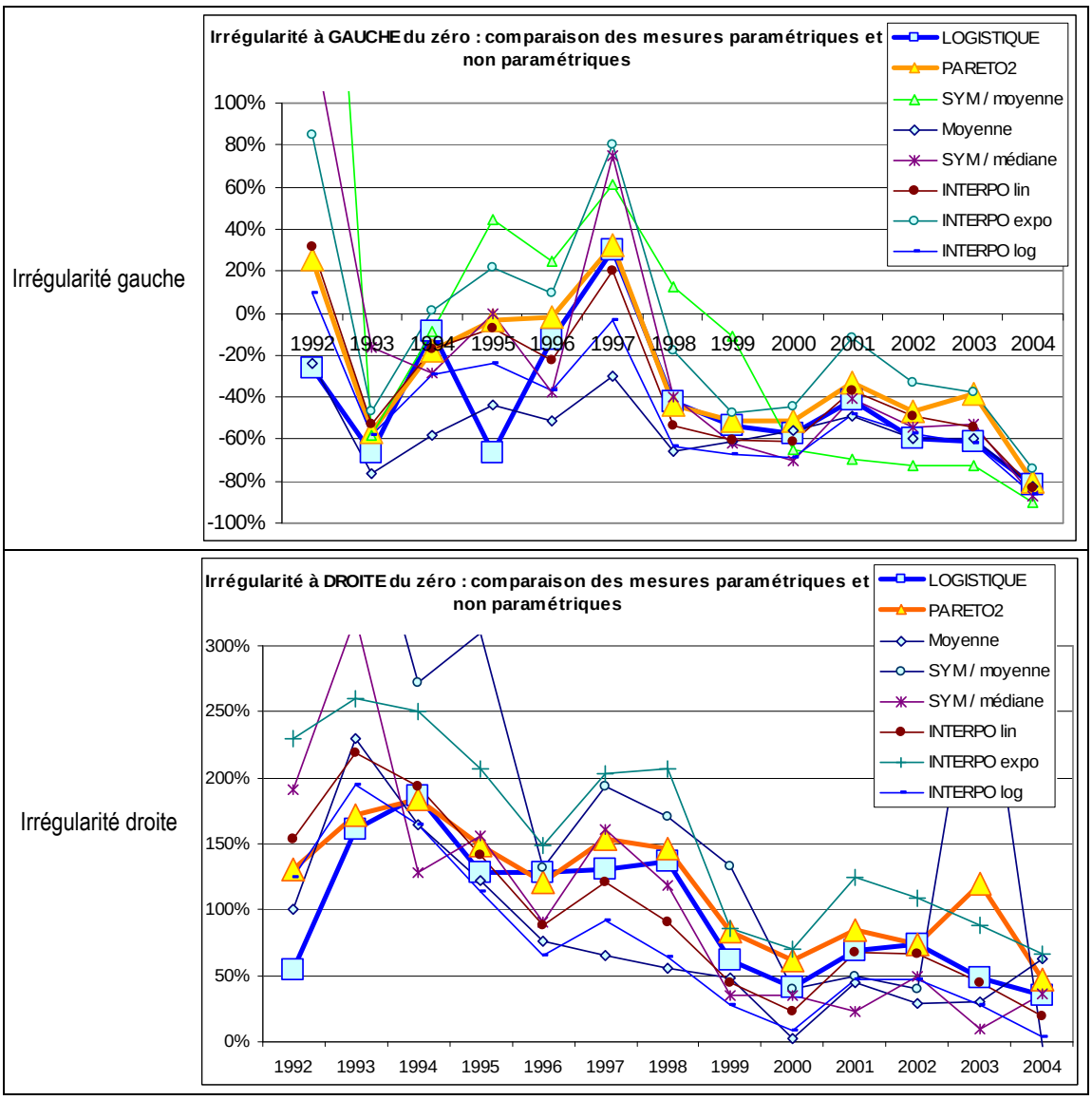
[174] Le coefficient de corrélation entre les deux séries d'irrégularités gauche est de 89%. Il passe à 97% si on retire les deux seules années (1992 et 1995) où des écarts sensibles sont visibles. Le coefficient de corrélation entre les deux séries d'irrégularités droite est de 87%. Il passe à 98% si on retire les deux seules années (1992 et 2003) où des écarts sensibles sont visibles.

positive à droite du seuil, confirmant ainsi les hypothèses issues de la littérature prévoyant des manipulations comptables pour éviter un résultat faiblement négatif.

b. Comparaison des méthodes paramétriques et non paramétriques

Visuellement, les mesures paramétriques apparaissent (Figure 121) comme étant « médianes » par rapport aux autres mesures. Ce sont de bonnes synthèses. Elles suivent les tendances, mais leur évolution est plus lissée.

Figure 121 : Comparaison des irrégularités mesurées par les méthodes paramétrique avec les irrégularités mesurées par des méthodes non paramétriques



Dans la partie 2, l'étude des mesures non paramétriques avait conduit à constater une certaine homogénéité des mesures (les coefficients de corrélations sont presque tous compris entre 80 et 92 %), mais le choix d'une méthode optimale était difficile. En l'absence d'argumentation théorique, les mesures par interpolation linéaire apparaissaient comme un bon compromis.

Tableau 95 : Corrélations entre les différentes mesures

Coefficients de corrélation ; irrégularité gauche								
	PARETO2	Logistiq	symétrie /moyenne	symétrie / médiane	interpolation linéaire	interpolation exponentielle	interpolation logarithmique	Moyenne des corrélation
PARETO2	1,000	0,868	0,759	0,873	0,948	0,964	0,908	**0,903**
Logistiq	0,868	1,000	0,706	0,742	0,802	0,849	0,758	0,818
symétrie /moyenne	0,759	0,706	1,000	0,715	0,711	0,712	0,699	0,757
symétrie / médiane	0,873	0,742	0,715	1,000	0,936	0,893	0,934	0,871
interpolation linéaire	0,948	0,802	0,711	0,936	1,000	0,949	0,990	**0,905**
interpolation exponentielle	0,964	0,849	0,712	0,893	0,949	1,000	0,897	0,895
interpolation logarithmique	0,908	0,758	0,699	0,934	0,990	0,897	1,000	0,884

Coefficients de corrélation ; irrégularité droite								
	PARETO2	Logistiq	symétrie /moyenne	symétrie / médiane	interpolation linéaire	interpolation exponentielle	interpolation logarithmique	Moyenne des corrélation
PARETO2	1,000	0,885	0,888	0,810	0,925	0,966	0,897	**0,910**
Logistiq	0,885	1,000	0,641	0,781	0,851	0,904	0,815	0,840
symétrie /moyenne	0,888	0,641	1,000	0,726	0,828	0,795	0,830	0,815
symétrie / médiane	0,810	0,781	0,726	1,000	0,892	0,876	0,887	0,853
interpolation linéaire	0,925	0,851	0,828	0,892	1,000	0,961	0,996	**0,922**
interpolation exponentielle	0,966	0,904	0,795	0,876	0,961	1,000	0,933	**0,919**
interpolation logarithmique	0,897	0,815	0,830	0,887	0,996	0,933	1,000	0,908

L'utilisation des méthodes paramétriques permet de sortir de cette difficulté. Non seulement leurs mesures confirment les tendances déjà observées (les coefficients de corrélation des mesures par l'ajustement Pareto2 sont parmi les plus élevés, 90 à 91 % en moyenne, alors que l'ajustement Logistique est sensiblement moins bien corrélé avec les autres mesures -82 à 84 % seulement-), mais ces méthodes bénéficient de qualités supplémentaires :

(1) Elles reposent sur un fondement théorique que n'ont pas les méthodes non paramétriques.

(2) Le calcul de l'effectif théorique utilisé pour mesurer l'irrégularité n'est pas influencé par le choix de la largeur des intervalles.

(3) L'effectif de la population étudiée n'influence pas les mesures. À ce sujet, il apparaît que les irrégularités les plus extrêmes mesurées par les méthodes non paramétriques apparaissent les années où les effectifs sont les plus faibles. Il n'est pas interdit de penser que la sensibilité de ces mesures à la taille de l'échantillon soit, au moins partiellement, responsable de ces importants écarts.

(4) Les méthodes paramétriques permettent de représenter graphiquement les irrégularités et de les comparer à toutes les autres irrégularités qui se manifestent sur l'ensemble de la distribution. L'intérêt de la représentation graphique peut être souligné : l'irrégularité n'est pas seulement mesurée en valeur absolue, mais en valeur relative (par rapport à l'ensemble des irrégularités de la distribution). Ainsi, les irrégularités à gauche qui sont souvent moins prononcées que les irrégularités à droite apparaissent cependant comme étant généralement les irrégularités négatives les plus importantes de toute la distribution. Les mesures non paramétriques sont souvent complétées par des tests statistiques pour tenter de mettre en évidence la significativité de l'irrégularité. Grâce aux représentations graphiques, ces tests deviennent inutiles dans la mesure où observer qu'une irrégularité est la plus importante sur l'ensemble de la distribution est plus parlant et tout aussi riche d'enseignement.

(5) Enfin, les méthodes paramétriques permettent de mettre en évidence de nouveaux indicateurs d'irrégularité : identification du nombre total d'entreprises manipulatrices et estimation des montants manipulés. C'est ce qui est développé dans la section suivante.

B. Deux nouveaux indicateurs d'irrégularité

Jusqu'alors, quelle que soit la méthode employée, mesurer l'irrégularité au seuil du résultat nul impliquait de déterminer une largeur d'intervalle fixe. Dans cette dernière section, les mesures paramétriques permettent de lever la contrainte de largeur d'intervalle. L'étude peut alors optimiser la mesure des irrégularités en faisant varier la largeur de l'intervalle d'observation.

Deux nouveaux indicateurs peuvent alors être calculés : (1) le nombre total d'entreprises manipulatrices, et (2) le montant manipulé. La sous-partie 1 présente la démarche de l'étude, la sous-partie 2 présente les calculs année par année, et la sous-partie 3 fait la synthèse des résultats.

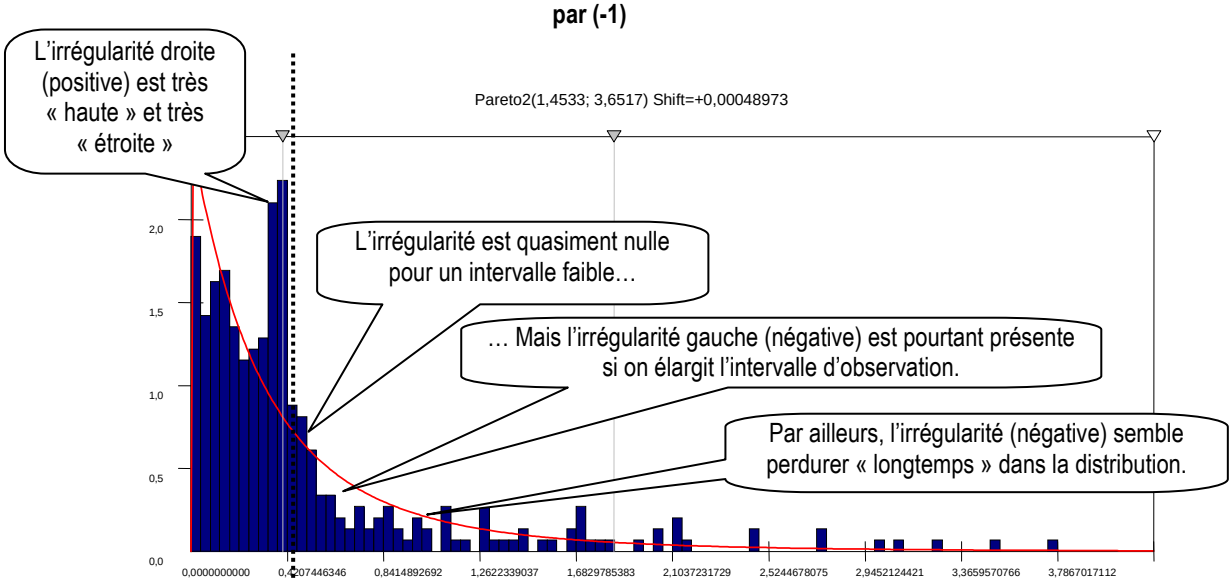
1. Démarche de l'étude

a. « Largeur » et « Hauteur » de l'irrégularité

La partie 2 de la thèse a montré que les irrégularités sont sensibles à la largeur d'intervalle utilisée pour représenter les distributions. L'observation graphique de la distribution des irrégularités, rendue possible grâce à une démarche paramétrique, conduit à constater que l'irrégularité à droite du seuil du résultat nul est généralement plus ramassée et plus prononcée (le « pic » est plus pointu) alors que l'irrégularité à gauche du seuil du résultat nul est généralement moins prononcée, mais plus « persistante » (le « trou » est peu « profond » mais plus « large »). Autrement dit, le sous-effectif des entreprises déficitaires demeure visible pour une largeur d'intervalle plus importante.

Figure 122 : Partie gauche de la distribution (centrée sur la médiane, réduite, et transposée) des résultats des entreprises française en 1997

L'irrégularité à droite du seuil du résultat nul apparaît à gauche du trait pointillé vertical sur le graphe puisque la distribution (des valeurs négatives centrées sur la médiane) est transformée en valeur positives en multipliant par (-1)



Ce constat conduit à imaginer une mesure de l'irrégularité qui ne se limite pas à une largeur d'observation fixe, mais qui au contraire cherche à compter toutes les entreprises irrégulières. Cette démarche conduit à mesurer :

- La largeur de l'irrégularité, c'est-à-dire la largeur maximale permettant de mettre en évidence une irrégularité significative.
- La hauteur de l'irrégularité, c'est-à-dire l'irrégularité maximale (sur ou sous-effectif maximum) constatée sur l'intervalle définit précédemment (la largeur de l'irrégularité).

b. Intérêt de ces deux nouveaux indicateurs

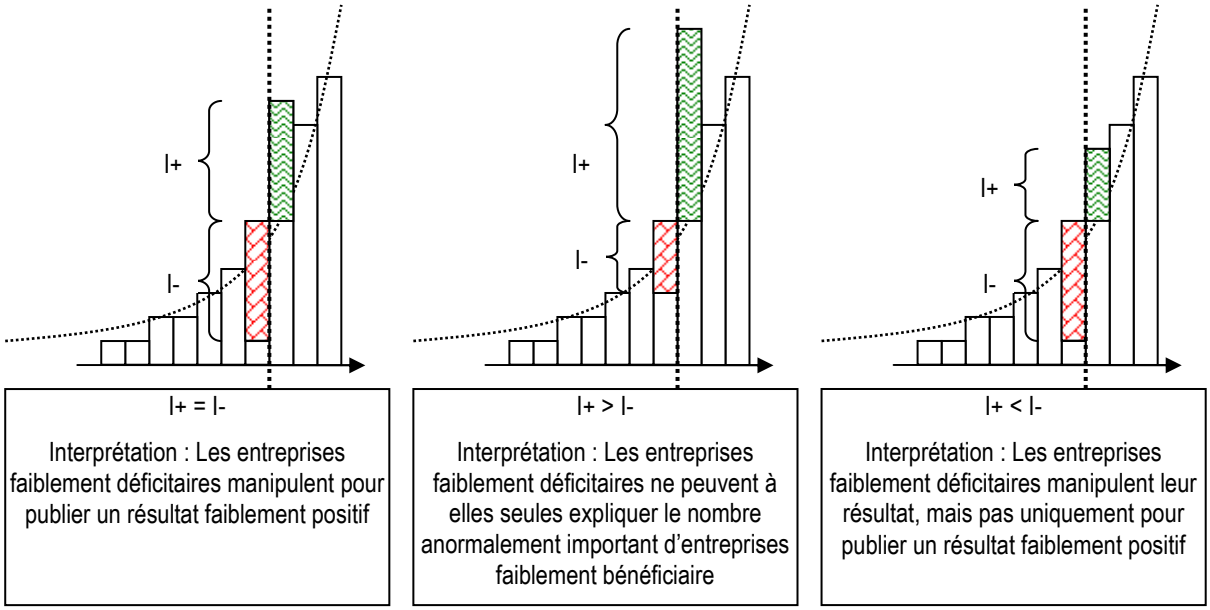
(1) Comme il est possible de déterminer le nombre total d'entreprises en sur ou sous-effectif, il devient possible de mesurer l'irrégularité, non plus en l'exprimant par une mesure « locale », mais en l'exprimant en fonction de l'ensemble de la population. Autrement dit, au lieu de dire qu'il y a, sur un intervalle donné, X % d'entreprises irrégulières, ou que l'irrégularité est sur un intervalle donné significative a Y %, il est possible de mesurer et d'affirmer que Z % de la population totale manipule ses comptes pour éviter de publier une perte.

(2) Comme l'irrégularité à droite du seuil du résultat nul est mesurée indépendamment de l'irrégularité à gauche, il est possible de comparer le nombre d'entreprises en sur-effectif (à droite) à celui des entreprises manquantes (à gauche). Dans la littérature, les auteurs présument que les manipulations comptables pour éviter le seuil du résultat nul sont de faible ampleur. Les entreprises faiblement déficitaires manquantes seraient donc celles qui se retrouvent en surnombre parmi les entreprises faiblement bénéficiaires (voir Figure 123). Autrement dit, l'irrégularité à droite doit être sensiblement la même (au signe près) que l'irrégularité à gauche. Dans le cas contraire :

- Si l'irrégularité à gauche est supérieure (en valeur absolue) à l'irrégularité à droite, cela signifie que les entreprises qui évitent de publier une petite perte ne se retrouvent pas toutes parmi les entreprises publiant un petit bénéfice.

- Si l'irrégularité à droite est supérieure (en valeur absolue) à l'irrégularité à gauche, cela signifie que les entreprises irrégulièrement nombreuses publiant un petit bénéfice ne sont pas toutes des entreprises qui auraient publié un résultat faiblement déficitaire sans manipulation.

Figure 123 : Différentes interprétations de la comparaison des irrégularités droite et gauche



(3) Enfin, il est possible d'évaluer les montants manipulés. Ceux-ci se déduisent de la largeur de l'irrégularité pour dépasser le seuil du résultat nul. Il est alors possible de distinguer (a) le montant maximal manipulé, et (b) le montant moyen manipulé par les entreprises qui évitent de publier une perte.

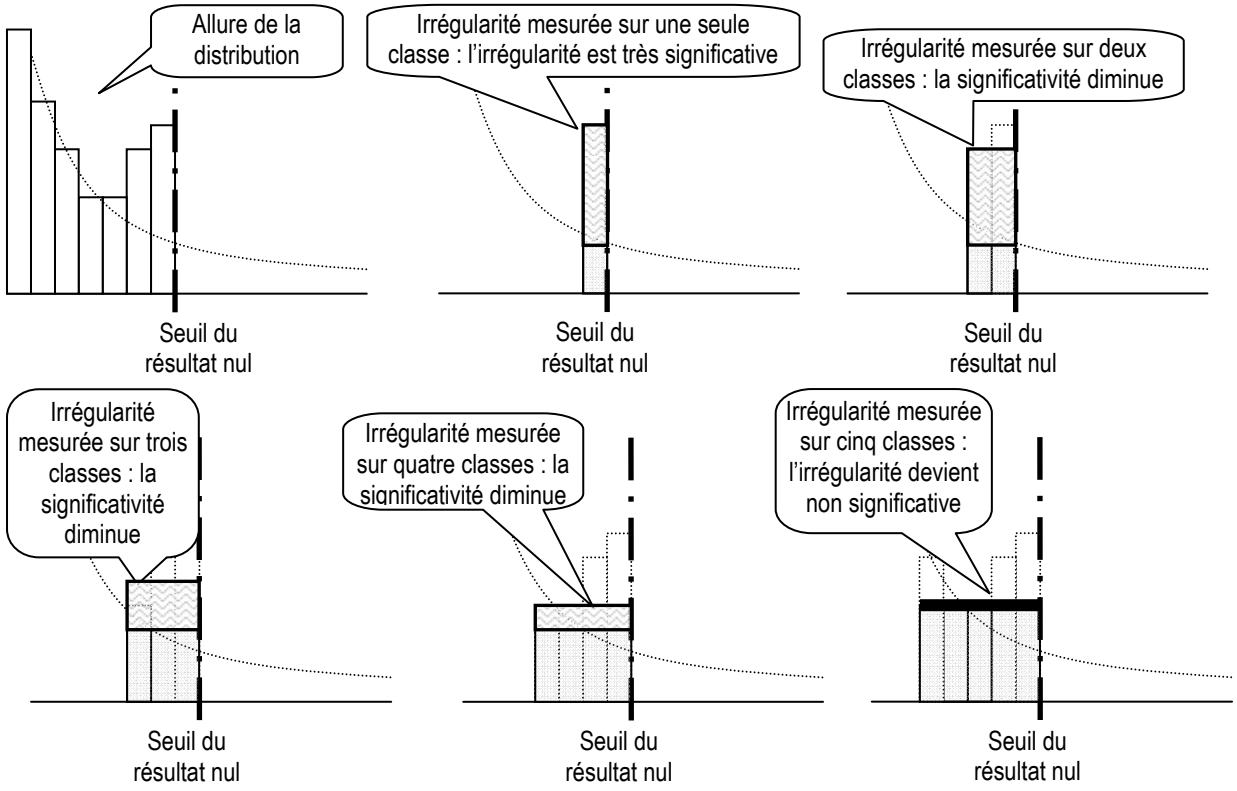
Toutes ces mesures nécessitent de mesurer, année après année, la largeur maximale permettant d'observer une irrégularité significative. Un seuil de significativité doit donc être préalablement établi.

c. Détermination du seuil de significativité

c.1. Démarche générale

Pour mesurer le nombre total d'entreprises manipulatrices, et les montants (maximum et moyen) manipulés, il faut dans un premier temps déterminer un seuil d'irrégularité minimum au dessus duquel l'irrégularité sera considérée significative. Chaque année, l'irrégularité est mesurée pour une largeur d'intervalle minimale de 0,05 (pour éviter les phénomènes aberrants qui peuvent être constatés lorsque les intervalles sont trop étroits). Puis la largeur est augmentée jusqu'à ce que l'irrégularité ne soit plus perceptible. La largeur maximale de l'irrégularité est atteinte lorsque l'irrégularité devient inférieure ou égale au seuil minimum déterminé.

Figure 124 : Recherche de la largeur maximale de l'intervalle irrégulier



Pour déterminer le seuil de significativité de l'irrégularité, deux démarches sont possibles. La première consiste à fixer arbitrairement un seuil à l'irrégularité relative. Par exemple, si l'on considère qu'une irrégularité de plus de 50 % est significative, cela signifie que la

moitié au moins des entreprises attendues seront en trop ou manquantes. La seconde consiste à effectuer un test statistique. La sous-partie suivante explique pourquoi cette seconde démarche n'est pas retenue.

c.2. Test statistique de significativité

Burgstahler et Dichev (1997) proposent un test statistique pour mesurer la significativité des irrégularités qu'ils observent. Ils considèrent que la distance entre le nombre observé et le nombre attendu d'entreprises sur un intervalle divisé par l'écart type (σ) de la différence suit une loi normale[175]. Pour estimer la variance de la distance, les auteurs estiment qu'elle doit être approximativement égale à la somme des variances, soit :

$\sigma = \sqrt{[N \times p_{obs} \times (1 - p_{obs}) + N \times p_{théo} \times (1 - p_{théo})]}$

Le test statistique est fondé sur l'hypothèse nulle H_0 : en cas d'absence de gestion de résultat, $p_{obs} = p_{théo}$

Plus précisément, les auteurs considèrent que la proportion, dans la population totale, des effectifs d'une classe i suit une loi binomiale (loi des tirages aléatoires avec remise) de paramètre $\beta(N, p_i)$ avec N = taille de la population et p_i = effectif de la classe exprimé en pourcentage de la population totale (ou probabilité qu'un résultat tiré au hasard dans la population se situe sur cette classe). Rappelons que la loi binomiale (qui est une somme de lois de Bernouilli de même paramètre p) de paramètres $\beta(N, p)$ a pour espérance Np et pour variance Np(1-p). Les auteurs estiment que le nombre d'observations situées sur une classe est une variable aléatoire approximativement indépendante des effectifs des intervalles adjacents. Ils appliquent donc une des propriétés mathématiques de la loi binomiale : la somme de deux variables aléatoires indépendantes de lois binomiales $\beta(N, p_1)$ et $\beta(N, p_2)$ est une variable aléatoire de loi $\beta(N, p_1+p_2)$. Il en découle que la variance de l'effectif observé sur la classe i = $N\, p_{i\ obs}\,(1-p_{i\ obs})$ et que la variance de l'effectif attendu (ou théorique) sur la classe i = $N\, pi_{\ théo}\,(1-p_{i\ théo})$.

Dans leur article, les auteurs estiment $p_{théo}$ par la moyenne des deux classes adjacentes. Donc $p_{i\ théo} = (p_{i+1} + p_{i-1}) / 2$. Donc la variance de l'effectif attendu = $N \times ¼ \times (p_{i+1} + p_{i-1}) \times (1 - p_{i+1} - p_{i-1})$. In fine, les auteurs considèrent qu'en absence de gestion du résultat, la différence standardisée D devrait suivre une loi normale centrée réduite, avec : $D = N(p_i - p_{i\ théo}) / \sqrt{[Np_i(1-p_i) + (1/4)\, N\, (p_{i-1} + p_{i+1})\, (1 - p_{i-1} - p_{i+1})]}$

[175] Ce postulat n'est cependant pas démontré.

Dans le cadre de l'étude, ce même test a été réalisé sur chacune des irrégularités annuelles mesurées. Mais l'effectif théorique n'étant pas estimé par moyenne des effectifs adjacents, la variance de la différence n'est pas égale à la somme des variances de deux variables indépendantes. Le test a donc été réalisé comme un test de proportion où la variable $(p_{obs} - p_{théo}) / \sigma$ suit une loi Normale, avec $\sigma = \sqrt{[p_{théo} x(1- p_{théo})/N]}$. Le Tableau 96 présente les résultats.

Tableau 96 : Significativité des irrégularités annuelles

	effectif théorique	effectif observé	écart	**écart relatif= (écart/eff.théo)**	N	pobs	pthéo	Abs(pobs-pthéo)	√[p(1-p)/n]	z	**Significativité**
	Irrégularité gauche										
1992	11.19	14	2.81	25.1%	291	0.048	0.038	0.010	0.0113	0.8567	-
1993	11.6	5	-6.6	**-56.9%**	296	0.017	0.039	0.022	0.0113	1.9770	**
1994	12.18	10	-2.18	-17.9%	376	0.027	0.032	0.006	0.0091	0.6350	-
1995	13.44	13	-0.44	-3.3%	422	0.031	0.032	0.001	0.0085	0.1220	-
1996	16.3	16	-0.3	-1.8%	616	0.026	0.026	0.000	0.0065	0.0753	-
1997	15.85	21	5.15	**32.5%**	702	0.030	0.023	0.007	0.0056	1.3084	*
1998	15.98	9	-6.98	**-43.7%**	753	0.012	0.021	0.009	0.0053	1.7649	**
1999	16.48	8	-8.48	**-51.5%**	797	0.010	0.021	0.011	0.0050	2.1108	**
2000	18.39	9	-9.39	**-51.1%**	791	0.011	0.023	0.012	0.0054	2.2156	**
2001	19.43	13	-6.43	**-33.1%**	734	0.018	0.026	0.009	0.0059	1.4784	*
2002	20.68	11	-9.68	**-46.8%**	693	0.016	0.030	0.014	0.0065	2.1611	**
2003	14.63	9	-5.63	**-38.5%**	657	0.014	0.022	0.009	0.0058	1.4886	*
2004	15.11	3	-12.11	**-80.1%**	609	0.005	0.025	0.020	0.0063	3.1548	***
	Irrégularité droite										
1992	13,85	32	18,15	**131,0%**	291	0,11	0,05	0,062	0,0125	4,9974	***
1993	13,96	38	24,04	**172,2%**	296	0,13	0,05	0,081	0,0123	6,5915	***
1994	14,45	41	26,55	**183,7%**	376	0,11	0,04	0,071	0,0099	7,1226	***
1995	16,46	41	24,54	**149,1%**	422	0,10	0,04	0,058	0,0094	6,1702	***
1996	19,43	43	23,57	**121,3%**	616	0,07	0,03	0,038	0,0070	5,4335	***
1997	18,54	47	28,46	**153,5%**	702	0,07	0,03	0,041	0,0061	6,6987	***
1998	18,63	46	27,37	**146,9%**	753	0,06	0,02	0,036	0,0057	6,4211	***
1999	19,13	35	15,87	**83,0%**	797	0,04	0,02	0,020	0,0054	3,6728	***
2000	21,7	35	13,3	**61,3%**	791	0,04	0,03	0,017	0,0058	2,8951	***
2001	22,69	42	19,31	**85,1%**	734	0,06	0,03	0,026	0,0064	4,1180	***
2002	24,12	42	17,88	**74,1%**	693	0,06	0,03	0,026	0,0070	3,7057	***
2003	15,48	34	18,52	**119,6%**	657	0,05	0,02	0,028	0,0059	4,7636	***
2004	17,65	26	8,35	**47,3%**	609	0,04	0,03	0,014	0,0068	2,0170	**

*** = significatif à 99% (z>2,33) ; ** = significatif à 95% (z>1,64) ; * = significatif à 90% (z>1,28)

Les irrégularités à gauche du seuil sont peu significatives avant 1997. Les irrégularités significatives à 95 % sont toutes supérieures à 43 % en valeur relative. Les irrégularités non significatives sont toutes inférieures à 25 % en valeur relative. Les irrégularités à droite du seuil apparaissent presque toutes significatives à 99 %. Une seule irrégularité n'est significative qu'à 95 %. Elle représente 47 % de l'effectif attendu en valeur relative.

Ces tests suggèrent donc que les irrégularités doivent être plus ou moins supérieures à 25 % pour être significatives.

Mais une telle conclusion ne peut pas être considérée comme définitive. Si Takeuchi (2004) montre que le test de discontinuité réalisé par Burgstahler et Dichev (1997) est performant quelle que soit la forme de la distribution (fonction linéaire, loi normale, loi du Chi-deux), le problème de ce test est qu'il est théoriquement impossible d'établir un seuil de significativité unique dans la mesure où la significativité d'une irrégularité dépend de la taille (N) de la population (cette taille change chaque année) d'une part, et d'autre part, de la probabilité ($p_{théo}$) qu'une entreprise soit située sur l'intervalle étudié, probabilité dépendant elle-même de la position du seuil dans la distribution (et donc des paramètres annuels de la distribution).

Autrement dit, toutes choses égales par ailleurs, plus N est grand, plus σ est petit, et plus z est grand. La même irrégularité relative peut donc être significative lorsque la taille de l'échantillon est importante, alors qu'elle peut ne pas l'être lorsque la taille est réduite. Par ailleurs, toutes choses égales par ailleurs, plus $p_{théo}$ est grand (irrégularité située proche du sommet, ou élargissement de la largeur d'intervalle), plus σ est petit, et donc moins l'irrégularité doit être importante pour être significative. Autrement dit, une même irrégularité relative apparaît plus significative les années où le nombre d'observations est important, et où l'irrégularité est située proche du sommet. Cela signifie également que les irrégularités à droite du seuil étant toujours plus proches du sommet que les irrégularités gauches, elles sont plus facilement considérées comme significatives que les irrégularités gauches.

Comme les tests dépendent des paramètres annuels des distributions, il faudrait établir un seuil de significativité ajusté à chaque distribution annuelle. Par ailleurs, ce seuil de significativité devrait être modifié en fonction de la largeur de l'intervalle étudié. Une telle démarche est extrêmement fastidieuse et peu pragmatique lorsque l'étude cherche à optimiser la mesure de l'irrégularité en modifiant la largeur de l'intervalle d'étude.

Par ailleurs, il n'est pas inutile de rappeler que le test de Burgstahler et Dichev (1997) repose sur quatre postulats discutables : (1) l'écart standardisé entre $p_{théo}$ et p_{obs} suit une loi Normale[176], (2) la variance est estimée par la somme des deux variances[177] de $p_{théo}$ et

[176] Hypothèse non démontrée.

[177] C'est une approximation.

p_{obs},(3) $p_{théo}$ est estimé par la moyenne des probabilités des intervalles adjacents[178], et (4) les effectifs des intervalles adjacents sont considérés comme indépendants[179].

Toutes ces limites rendent l'utilisation d'un test de significativité délicat à mettre en œuvre pour déterminer un seuil de significativité opérationnel. C'est la raison pour laquelle elle est abandonnée au profit d'un seuil arbitraire unique exprimé en fonction de l'irrégularité relative.

c.3. Détermination empirique d'un seuil de significativité

À partir de quand une irrégularité (exprimée en valeur relative par souci de pragmatisme) doit-elle être considérée comme significative ? Pour déterminer cette valeur minimum, trois critères ont été retenus :

(1) Tout d'abord, une simple intuition permet de penser qu'une irrégularité de l'ordre de 50 % est importante. Une irrégularité de 100 % semble largement significative. En revanche, une irrégularité de 10 ou 20 % semble intuitivement peu significative. Autrement dit, la seule intuition conduit à penser que l'irrégularité minimum doit être située dans un ordre de grandeur de 20 à 50 %.

(2) Par ailleurs, les tests réalisés précédemment (Tableau 96) ont montré que les irrégularités significatives sont toujours (en valeur relative) supérieures à 32,5 % dans le cadre de notre population d'entreprises cotées françaises.

(3) Enfin, puisque la significativité dépend de N et de $p_{théo}$, et que ces deux valeurs varient d'une année à l'autre, des tests de proportion ont été réalisés chaque année afin de déterminer la valeur de l'irrégularité minimum annuelle significative à 99 %, 95 % et 90 %. Les résultats de ces calculs sont présentés dans le Tableau 97.

[178] Les limites de cette estimation sont développées dans la partie 2 de la thèse.

[179] Ce postulat est d'autant plus discutable qu'il y a une présomption de manipulation, donc de transfert d'un intervalle vers un autre.

Tableau 97 : Irrégularité annuelle minimum pour être significative

	effectif théorique	effectif observé	écart	**écart relatif**	N	z		Significativité à 90%				Significativité à 95%				Significativité à 99%			
								si z=	alors p_{obs}=	soit	soit	**si z=**	alors p_{obs}=	soit	soit	**si z=**	alors p_{obs}=	soit	soit
Irrégularité gauche																			
1992	**11.19**	**14**	**2.81**	**25.1%**	**291**	**0.857**	**-**	**1.28**	**0.024**	**7.0**	**37.52%**	**1.64**	**0.020**	**5.8**	**48.07%**	**2.58**	**0.009**	**2.7**	**75.63%**
1993	11.6	5	-6.6	-56.9%	296	1.977	**	1.28	0.025	7.3	36.84%	1.64	0.021	6.1	47.20%	2.58	0.010	3.0	74.25%
1994	12.18	10	-2.18	-17.9%	376	0.635	-	1.28	0.021	7.8	36.08%	1.64	0.017	6.5	46.22%	2.58	0.009	3.3	72.72%
1995	13.44	13	-0.44	-3.3%	422	0.122	-	1.28	0.021	8.8	34.35%	1.64	0.018	7.5	44.02%	2.58	0.010	4.1	69.25%
1996	16.3	16	-0.3	-1.8%	616	0.075	-	1.28	0.018	11.2	31.28%	1.64	0.016	9.8	40.08%	2.58	0.010	6.0	63.05%
1997	15.85	21	5.15	32.5%	702	1.308	*	1.28	0.015	10.8	31.79%	1.64	0.013	9.4	40.73%	2.58	0.008	5.7	64.07%
1998	**15.98**	**9**	**-6.98**	**-43.7%**	**753**	**1.765**	**	**1.28**	**0.014**	**10.9**	**31.68%**	**1.64**	**0.013**	**9.5**	**40.59%**	**2.58**	**0.008**	**5.8**	**63.85%**
1999	**16.48**	**8**	**-8.48**	**-51.5%**	**797**	**2.111**	**	**1.28**	**0.014**	**11.3**	**31.20%**	**1.64**	**0.012**	**9.9**	**39.98%**	**2.58**	**0.008**	**6.1**	**62.89%**
2000	18.39	9	-9.39	-51.1%	791	2.216	**	1.28	0.016	13.0	29.50%	1.64	0.014	11.4	37.80%	2.58	0.009	7.5	59.46%
2001	19.43	13	-6.43	-33.1%	734	1.478	*	1.28	0.019	13.9	28.65%	1.64	0.017	12.3	36.71%	2.58	0.011	8.2	57.75%
2002	20.68	11	-9.68	-46.8%	693	2.161	**	1.28	0.022	14.9	27.72%	1.64	0.019	13.3	35.52%	2.58	0.013	9.1	55.88%
2003	14.63	9	-5.63	-38.5%	657	1.489	*	1.28	0.015	9.8	33.09%	1.64	0.013	8.4	42.40%	2.58	0.007	4.9	66.70%
2004	15.11	3	-12.11	-80.1%	609	3.155	***	1.28	0.017	10.2	32.52%	1.64	0.014	8.8	41.66%	2.58	0.009	5.2	65.54%
										moyenne	32.48%				41.61%				65.46%
										max	**37,52%**				**48,07%**				**75,63%**
Irrégularité droite																			
1992	**13,85**	**32**	**18,15**	**131,0%**	**291**	**4,997**	***	**1,28**	**0,064**	**18,5**	**33,57%**	**1,64**	**0,068**	**19,8**	**43,01%**	**2,58**	**0,080**	**23,2**	**67,66%**
1993	13,96	38	24,04	172,2%	296	6,592	***	1,28	0,063	18,6	33,44%	1,64	0,067	19,9	42,85%	2,58	0,079	23,4	67,40%
1994	14,45	41	26,55	183,7%	376	7,123	***	1,28	0,051	19,2	33,02%	1,64	0,055	20,6	42,31%	2,58	0,064	24,1	66,55%
1995	16,46	41	24,54	149,1%	422	6,170	***	1,28	0,051	21,6	30,93%	1,64	0,054	23,0	39,63%	2,58	0,063	26,7	62,34%
1996	19,43	43	23,57	121,3%	616	5,433	***	1,28	0,041	25,0	28,58%	1,64	0,043	26,5	36,61%	2,58	0,050	30,6	57,60%
1997	18,54	47	28,46	153,5%	702	6,699	***	1,28	0,034	24,0	29,33%	1,64	0,036	25,5	37,58%	2,58	0,042	29,5	59,12%
1998	**18,63**	**46**	**27,37**	**146,9%**	**753**	**6,421**	***	**1,28**	**0,032**	**24,1**	**29,29%**	**1,64**	**0,034**	**25,6**	**37,52%**	**2,58**	**0,039**	**29,6**	**59,03%**
1999	**19,13**	**35**	**15,87**	**83,0%**	**797**	**3,673**	***	**1,28**	**0,031**	**24,7**	**28,91%**	**1,64**	**0,033**	**26,2**	**37,04%**	**2,58**	**0,038**	**30,3**	**58,28%**
2000	21,7	35	13,3	61,3%	791	2,895	***	1,28	0,035	27,6	27,10%	1,64	0,037	29,2	34,72%	2,58	0,042	33,6	54,62%
2001	22,69	42	19,31	85,1%	734	4,118	***	1,28	0,039	28,7	26,45%	1,64	0,041	30,4	33,89%	2,58	0,047	34,8	53,32%
2002	24,12	42	17,88	74,1%	693	3,706	***	1,28	0,044	30,3	25,61%	1,64	0,046	32,0	32,81%	2,58	0,053	36,6	51,61%
2003	15,48	34	18,52	119,6%	657	4,764	***	1,28	0,031	20,5	32,15%	1,64	0,033	21,9	41,19%	2,58	0,039	25,5	64,80%
2004	17,65	26	8,35	47,3%	609	2,017	**	1,28	0,038	22,9	30,02%	1,64	0,040	24,4	38,47%	2,58	0,047	28,3	60,51%
										Moyenne	29,88%				38,28%				60,22%
										Max	**33,57%**				**43,01%**				**67,66%**

La lecture de ces résultats confirme que les années (1992 et 1993) où les effectifs sont les plus réduits, ou lorsque la probabilité théorique est la plus faible (1998 et 1999), les irrégularités relatives doivent être, toutes choses égales par ailleurs, relativement élevées pour être jugées significatives. Ce tableau permet cependant de se forger un ordre de grandeur approximatif sur la valeur relative de l'irrégularité minimale à retenir. Il apparaît qu'une irrégularité de l'ordre de 65 % sera dans tous les cas jugée significative par un test de proportion avec moins d'1 % de chances de se tromper. Une irrégularité de l'ordre de 40 % sera dans tous les cas jugée significative avec moins de 5 % de chances de se

tromper. Enfin, une irrégularité de l'ordre de 30 % sera dans tous les cas jugée significative avec moins de 10 % de chances de se tromper.

Ces résultats conduisent donc à fixer la valeur de 40 % comme seuil de significativité[180] pour la suite de l'étude. Cette valeur garantit une significativité de l'irrégularité à plus de 95 % quelle que soit l'année étudiée, et ce même lorsque la taille de l'échantillon est la plus faible ou que l'irrégularité est située loin du sommet. Ce seuil apparaît comme raisonnablement prudent[181].

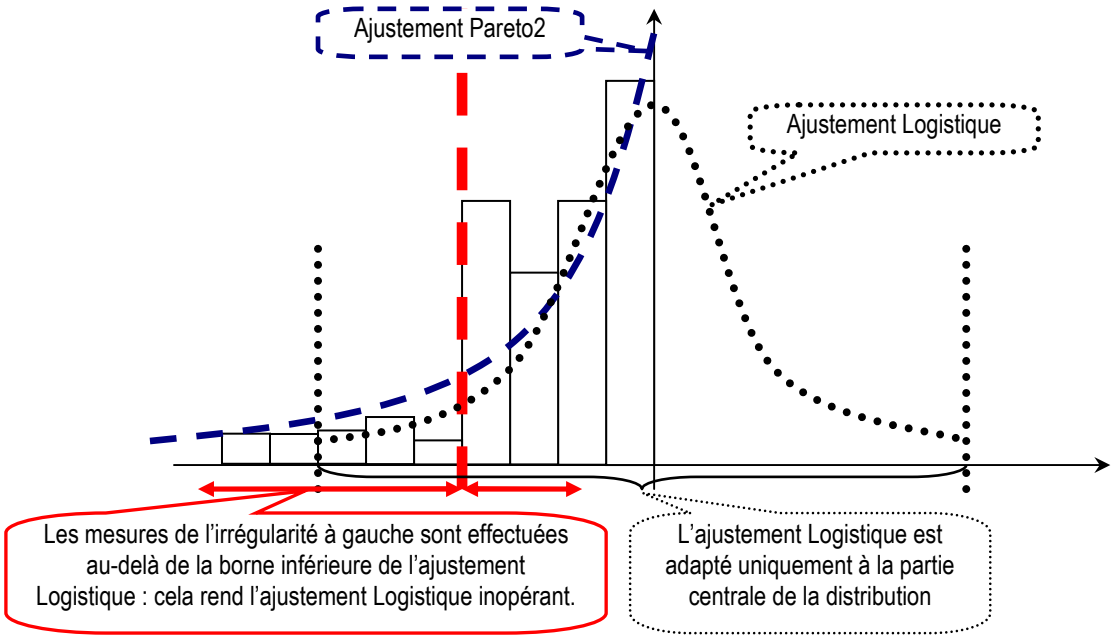
2. Mesures année par année

Cette section détaille les mesures d'irrégularités effectuées en faisant varier la largeur des intervalles situés de part et d'autre du seuil résultat nul. Les irrégularités sont mesurées par la différence entre la distribution réelle et une distribution théorique suivant une loi Pareto2. Les irrégularités mesurées selon l'ajustement logistique ne sont pas opérationnelles dans le cadre de cette partie de l'étude. En effet, si la loi Logistique peut ajuster correctement la partie centrale de la distribution, elle ne permet pas d'effectuer des mesures en élargissant l'intervalle d'étude au-delà de la partie centrale (voir Figure 125). C'est pourquoi seul l'ajustement Pareto2 est utilisé.

[180] Le seuil unique recherché a l'avantage d'être facilement opérationnel. Il faut cependant souligner qu'il « pénalise » doublement les irrégularités à gauche du seuil. En effet, puisque l'irrégularité est toujours située dans la partie gauche de la distribution, l'irrégularité gauche est toujours plus éloignée du sommet, et $p_{théo}$ est toujours inférieur à gauche qu'il ne l'est à droite. Donc la même irrégularité (pour la même année) paraît toujours plus significative si elle est considérée comme une irrégularité à droite. Cette limite apparaît d'autant plus importante lorsque la largeur de l'intervalle est agrandie. Par ailleurs, il faut souligner que l'irrégularité à gauche étant négative, elle ne peut être supérieure à 100% (il ne peut y avoir plus d'entreprises manquantes que d'entreprises attendues). Les irrégularités droites ne connaissent pour leur part aucune limite maximum.

[181] Afin de contrôler les résultats menés par la suite, les calculs sont également réalisés en abaissant le seuil à 30%. Les résultats ne sont pas fondamentalement différents.

Figure 125 : Pourquoi l'ajustement Logistique n'est pas retenu pour mesurer les irrégularités maximales



La largeur maximale (sous-partie a) correspond à la plus grande largeur permettant de révéler une irrégularité significative (c'est-à-dire > 40 %) à gauche comme à droite. Une fois la largeur maximale déterminée, trois indicateurs sont calculés. La sous-partie b présente la mesure du nombre total d'entreprises manipulatrices, les sous-parties c et d les montants maximaux et moyens manipulés.

a. Largeur maximale année par année

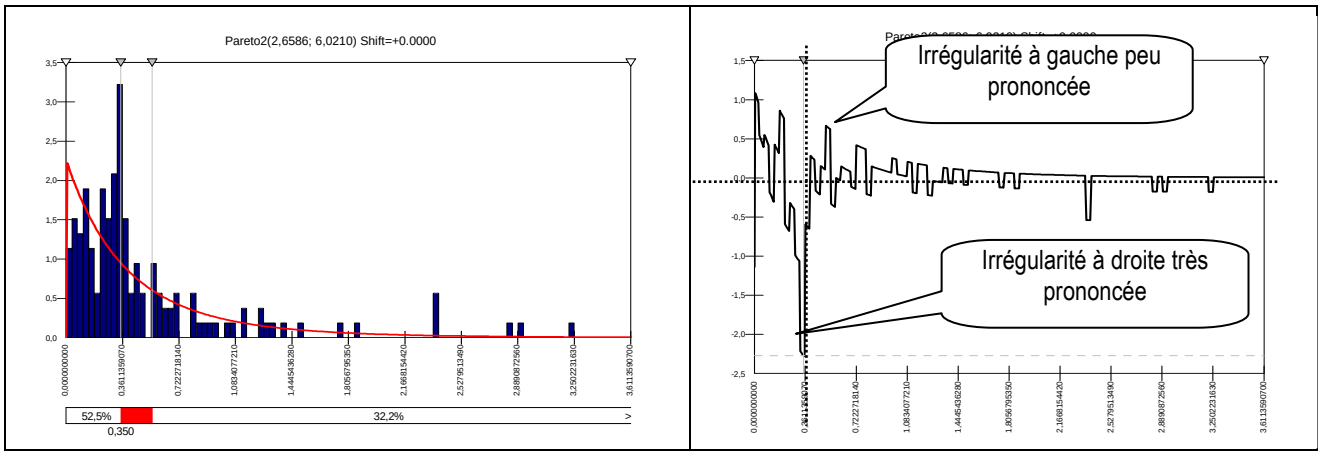
a.1. La démarche générale

Les mesures effectuées en 1992 sont détaillées pour expliciter la démarche suivie, puis les mesures effectuées de 1993 à 2004 sont détaillées en Annexe 14 : Mesure de la largeur maximale année par année. Le Tableau 100 synthétise les résultats.

a.2. Largeur maximale en 1992

Pour illustrer la démarche en 1992, la Figure 126 rappelle le graphique des irrégularités observées en 1992.

Figure 126 : Largeur maximale de l'intervalle irrégulier en 1992



Visuellement, l'irrégularité à gauche du seuil est peu prononcée. Ce résultat est confirmé par les mesures effectuées (Tableau 98). Même si elle est négative de manière très persistante, elle ne dépasse jamais -17 %, et demeure donc très en deçà du seuil de significativité (-40 %) retenu. En 1992, une valeur nulle est donc attribuée à la largeur de l'irrégularité (et par conséquent la hauteur maximale) à gauche du seuil.

Tableau 98 : Irrégularité gauche (en 1992) mesurée en élargissant l'intervalle d'étude

Bornes de l'intervalle	Distribution du résultat	De	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,04	-0,05	-0,06	-0,07	-0,1
		à	0	0	0	0	0	0	0	0	0	0	0
	Distribution ajustée	Soit de	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114
		à	0,45292	0,54471	0,63649	0,72828	0,82007	0,91185	1,09543	1,27899	1,46257	1,64614	2,19686
		Observé	14	19	27	34	35	39	43	47	50	51	53
		Ajusté	11,19	20,28	27,71	33,82	38,87	43,07	49,51	54,08	57,37	59,78	63,92
		écart	2,81	-1,28	-0,71	0,18	-3,87	-4,07	-6,51	-7,08	-7,37	-8,78	-10,92
		relatif	25%	**-6%**	**-3%**	1%	**-10%**	**-9%**	**-13%**	**-13%**	**-13%**	**-15%**	**-17%**

L'irrégularité à droite (voir Tableau 99) du seuil résultat nul est supérieure à 40 % jusqu'à une largeur maximale de 0,012. Le résultat s'élève alors à presque[182] 1,2 % de l'actif total. Sur cet intervalle, la distribution théorique prévoit 39,18 entreprises, alors que la distribution réelle en présente 56. Autrement dit, 16,8 entreprises sont en surnombre, soit 5,77 % du total des 291 entreprises présentes dans la population en 1992.

[182] Plus précisément, puisque nous étudions la distribution de la variable Résultat (IB) mise à l'échelle par (Actif Total – Résultat), si IB/(AT-IB)=0,012 aloIBs IB=(0,012/1,012) x AT soit IB=0,01186 x AT.

Tableau 99 : Irrégularité droite (en 1992) mesurée en élargissant l'intervalle d'étude

Bornes de l'intervalle	Distribution du résultat	De…	0	0	0	**0**	0	0	0	**0**	0	0	0
		À …	0,001	0,005	0,006	**0,007**	0,008	0,01	0,011	**0,012**	0,013	0,014	0,015
	Distribution ajustée	Soit de	0,34278	0,26934	0,25099	0,23263	0,21427	0,17756	0,15920	0,14084	0,12249	0,10413	0,08577
		à	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114	0,36114
		Observé	7	32	38	44	46	50	52	56	58	66	67
		Ajusté	2,54	13,86	17,01	20,3	23,74	31,11	35,05	39,18	43,5	48,03	52,78
		écart	4,46	18,14	20,99	**23,7**	22,26	18,89	16,95	**16,82**	14,5	17,97	14,22
		relatif	**176%**	**131%**	**123%**	**117%**	**94%**	**61%**	**48%**	**43%**	33%	37%	27%

La hauteur maximum est atteinte pour une largeur de 0,007 avec un surnombre de 23,7 entreprises, soit 23,7 / 291 = 8,14 % de l'échantillon. Elle constitue « l'irrégularité maximum » à droite du seuil.

Ce résultat appelle le commentaire suivant : le surnombre (8 % de l'échantillon) à droite du seuil n'est pas compensé par un sous-effectif à gauche du seuil. Autrement dit, la présomption de manipulation comptable de petite envergure, permettant à des entreprises qui ratent de peu le seuil, de publier malgré tout un résultat conforme aux attentes, n'est pas confirmée en 1992. Si les 23 à 24 entreprises, en surnombre au dessus du seuil du résultat nul, manipulent leur résultat pour dépasser le seuil, alors rien ne permet d'affirmer que le montant manipulé est faible.

a.3. Synthèse des mesures de largeurs max

Les largeurs maximales permettant de constater une irrégularité significative (supérieure à 40 % en valeur absolue) à gauche et à droite du seuil du résultat nul sont présentées dans le Tableau 100.

Tableau 100 : Largeur maximale permettant de constater une irrégularité significative (> 40 %)

Irrégularité gauche													
Irreg à gauche	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Largeur max	x	0,012	x	x	x	x	0,073	0,022	0,026	0,061	0,09	x	0,061
Irrégularité en nombre d'entreprises		11,73					43,71	23,14	29,85	55,12	15,19		39,36
Irrégularité droite													
Irrégularité à droite	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Largeur max	0,012	0,014	0,019	0,016	0,020	0,022	0,028	0,024	0,019	0,014	0,015	>=0,01845	0,017
Irrégularité en nombre d'entreprises	16,82	19,88	31,65	28,9	41,76	45,15	63	50,71	47,11	34,38	39,54	>=76,26	31,35

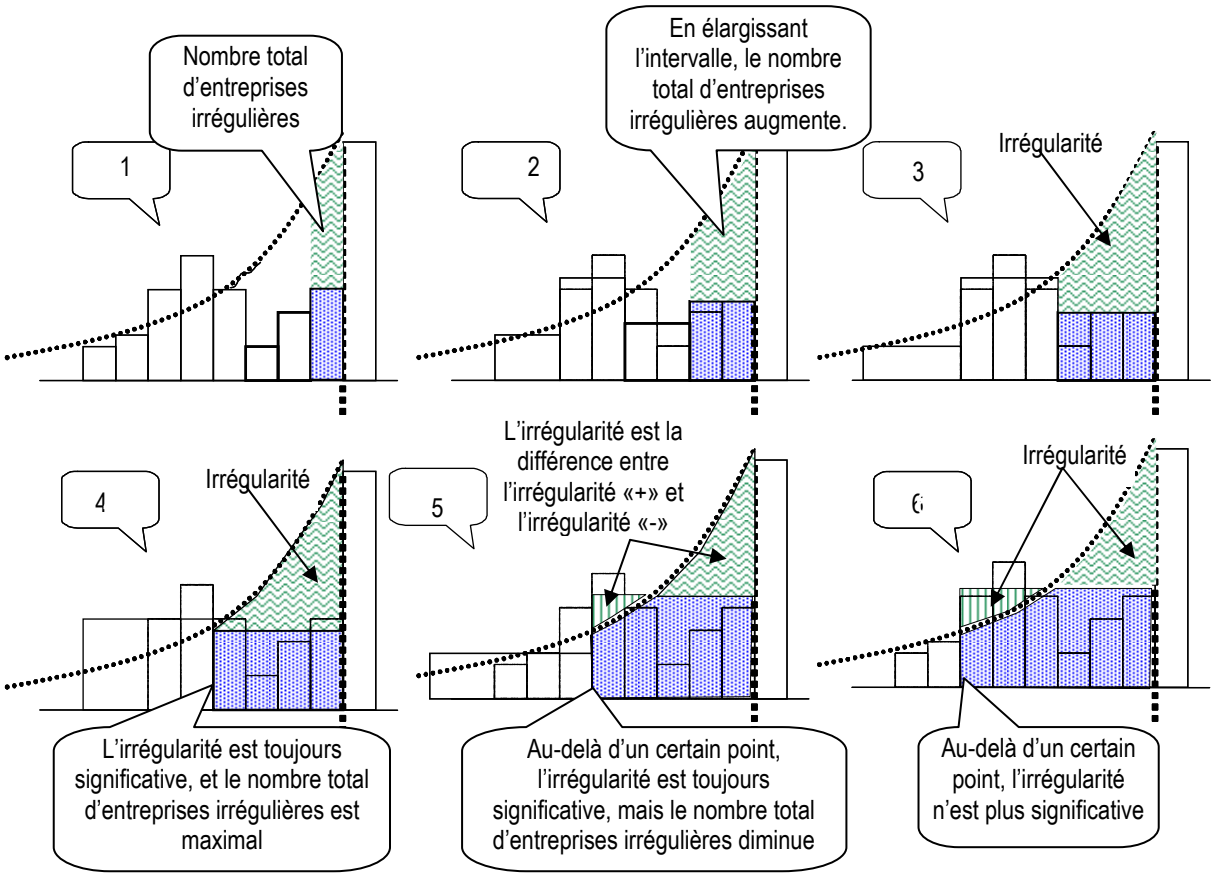
Il apparaît à la lecture de ce tableau que les irrégularités gauches sont généralement moins prononcées, donc souvent non significatives, mais lorsqu'elles sont significatives,

elles sont généralement plus « persistantes », c'est-à-dire visibles sur un intervalle plus large que les irrégularités droites. Ces résultats confirment le constat déjà réalisé visuellement. Cette analyse va être approfondie en déterminant, année après année, le nombre total d'entreprises manipulatrice, et les montants maximaux et moyens manipulés pour éviter de publier une perte.

b. Nombre total d'entreprises manipulatrices

Dans l'intervalle [0,005 ; largeur maximale], la plus grande « hauteur » de l'irrégularité est égale au nombre le plus élevé (non arrondi) d'entreprises irrégulières (cas n° 4 dans l'exemple de la Figure 127 ci-dessous). Cette valeur est exprimée de manière relative en la divisant par la population étudiée.

Figure 127 : Comment identifier le nombre total d'entreprises manipulatrices ?



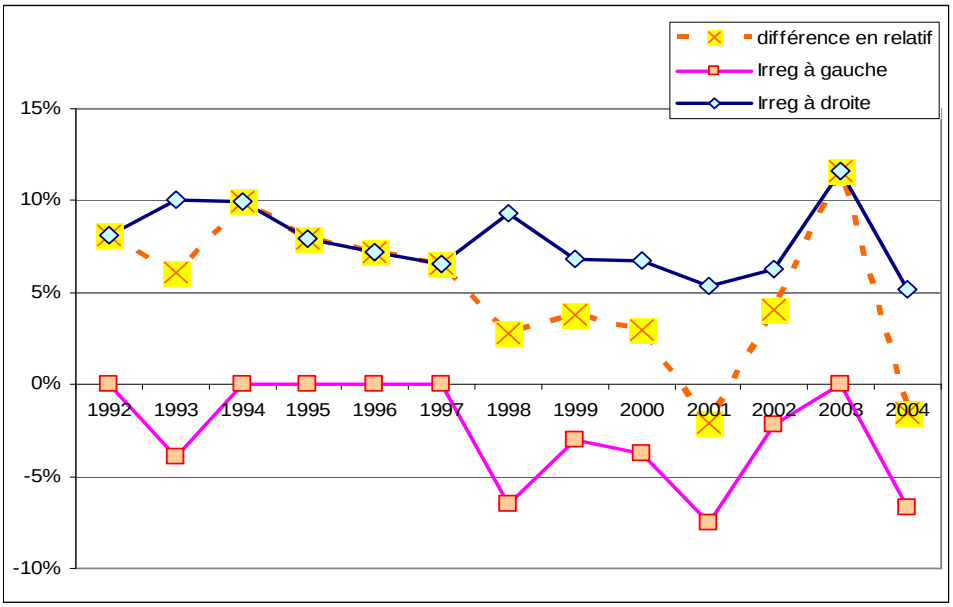
Autrement dit, le nombre d'entreprises irrégulières est maximisé. Le Tableau 101 récapitule ces mesures qui découlent des observations effectuées précédemment (section B-2. Mesures année par année).

Tableau 101 : Irrégularités annuelles maximales

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Irrégularité à gauche													
Largeur correspondante	*(0.1)*	0.012	*(0.07)*	*(0.07)*	*(0.1)*	*(0.1)*	0.06	0.02	0.026	0.061	0.009	*(0.1)*	0.06
Nombre d'entreprises max	*(-10.92)*	-11.73	*(-19.05)*	*(-19.74)*	*(-25.08)*	*(-28.72)*	-49.27	-24.15	-29.85	-55.12	-15.19	*(-73.6)*	-40.57
Nombre d'entreprises	291	296	*376*	*422*	*616*	*702*	753	797	791	734	693	*657*	609
Mesure relative	*(-3.8%)*	**-3.96%**	*(-5.1%)*	*(-4.7%)*	*(-4.1%)*	*(-4.1%)*	**-6.54%**	**-3.03%**	**-3.77%**	**-7.51%**	**-2.19%**	*(-11.2%)*	**-6.66%**
	(Les irrégularités entre parenthèse sont non significatives (<40%) ; elles ne sont pas représentées dans la Figure 128)												
Irrégularité à droite													
Largeur correspondante	0,007	0,007	0,013	0,013	0,014	0,013	0,020	0,023	0,013	0,012	0,010	>=0,01845	0,017
Nombre d'entreprises max	23.7	29.65	37.51	33.62	44.31	45.69	70	54.45	53.4	39.41	43.45	>=76.26	31.35
Nombre d'entreprises	291	296	376	422	616	702	753	797	791	734	693	657	609
Mesure relative	**8.14%**	**10.02%**	**9.98%**	**7.97%**	**7.19%**	**6.51%**	**9.30%**	**6.83%**	**6.75%**	**5.37%**	**6.27%**	**11.61%**	**5.15%**

Le nombre d'entreprises irrégulières à droite et à gauche du seuil est ensuite représenté graphiquement (Figure 128). La différence entre le sur-effectif à droite et le sous-effectif à gauche est également représentée.

Figure 128 : Représentation graphique des largeurs maximales à droite, à gauche, et leur différence



Ces résultats appellent les commentaires suivants :

(1) Confirmant la théorie de l'existence de manipulation comptable pour éviter le seuil du résultat nul, il apparaît que l'irrégularité à droite du seuil est toujours positive, alors que l'irrégularité à gauche du seuil, quand elle est significative, est toujours négative.

(2) Il apparaît une relative stabilité de l'irrégularité à droite. Chaque année, 5 à 10 % de la population publie un résultat anormalement faiblement positif. Ce surnombre tend à diminuer au cours de la période étudiée.

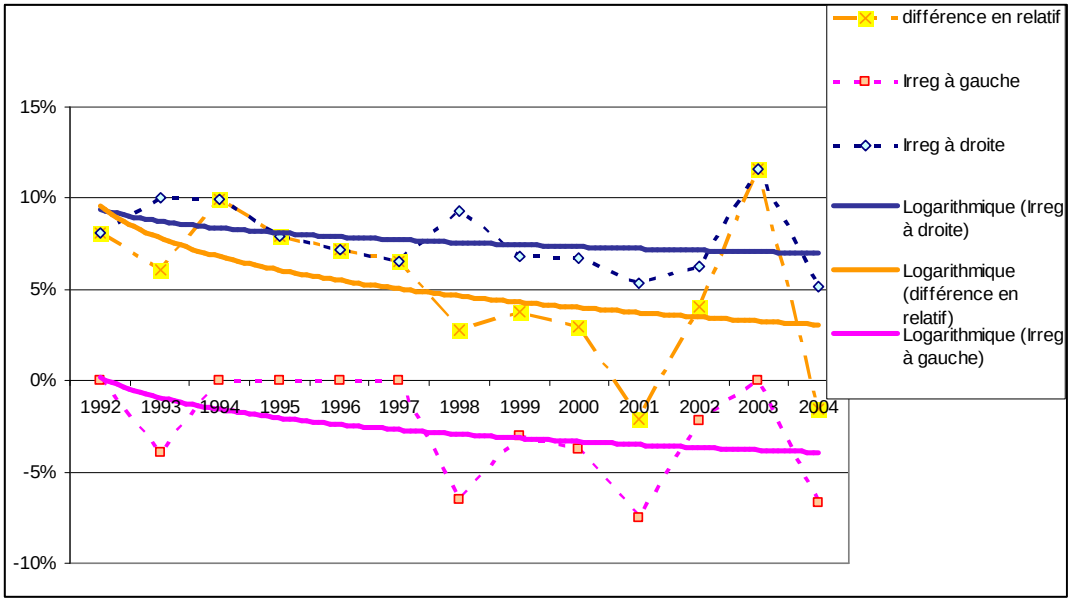
(3) En comparant ces mesures à celles (paramétriques ou non) effectuées en fixant une largeur d'intervalle, il apparaît que ces dernières accentuent les tendances. L'impact méthodologique du choix de la largeur entraîne des variations amplifiées. Ainsi, les irrégularités à droite, par exemple, oscillent entre 250 et 50 % durant la période considérée. Si la tendance (à la baisse) est confirmée, la mesure de l'irrégularité maximale exprimée en pourcentage de la population paraît beaucoup plus facile à interpréter qu'une mesure locale.

(4) L'irrégularité à gauche, c'est-à-dire le nombre d'entreprises faiblement déficitaires manquantes, connaît une évolution plus délicate à analyser. Deux phases sont observables :

- Jusqu'en 1997 (à l'exception de 1993), le nombre d'entreprises manquantes à gauche est non significatif. Cela signifie que l'hypothèse selon laquelle les entreprises cherchent à éviter de publier un résultat faiblement négatif n'est pas vérifiée. Autrement dit, la manipulation comptable pour dépasser le seuil du résultat nul n'est pas limitée à des toilettages des comptes marginaux.

- À partir de 1998, le nombre d'entreprises faiblement déficitaires devient significativement anormalement faible. L'hypothèse de petite manipulation pour éviter le seuil ne peut alors pas être écartée. Elle devient une explication possible.

Figure 129 : Courbes de tendance (logarithmique) de l'évolution des irrégularités maximales



(5) Durant toute la période étudiée, à l'exception de 2001 et 2004, les irrégularités à gauche sont toujours inférieures (en valeur absolue) aux irrégularités à droite. Autrement dit, si l'hypothèse de manipulation pour dépasser le seuil du résultat nul peut concerner une part assez constante de la population, l'hypothèse selon laquelle cette pratique ne concerne que des entreprises faiblement déficitaires n'est pas vérifiée, et ce, même après 1997.

c. Montant maximal manipulé

Il apparaît que, chaque année, l'irrégularité à droite est plus prononcée mais moins persistante (plus haute et moins large) que l'irrégularité à gauche. Il en découle que le sur-effectif à droite n'est pas compensé par le sous-effectif à gauche. La question qui surgit alors naturellement est : jusqu'où doit-on chercher les irrégularités à gauche pour compenser l'irrégularité à droite ? Autrement dit, quel sont les montants maximum manipulés pour éviter le seuil du résultat nul ? La démarche paramétrique permet de répondre à cette question puisqu'il est possible mesurer la largeur de la distribution à gauche du seuil nécessaire pour que le nombre d'entreprises manquantes compense le sur-effectif à droite. Dans ce contexte, les mesures sont effectuées sans se préoccuper de savoir si l'irrégularité à gauche est significative ou non. La démarche suppose deux conditions qui sont remplies sur les 13 années de l'étude :

- L'irrégularité à droite est positive (il y a un sur-effectif à droite).
- L'irrégularité à gauche est négative (les entreprises en sur-effectif à droite peuvent provenir des classes d'effectifs situés à gauche du seuil).

Le Tableau 102 rappelle le nombre d'entreprises en surnombre à droite du seuil, donc présumées avoir manipulé leurs comptes pour publier un petit bénéfice et le compare au nombre d'entreprises manquantes à gauche du seuil. Le montant maximal manipulé est égal à la somme de la largeur à gauche et de la largeur à droite (comme l'illustre la Figure 130).

Figure 130 : Illustration de la démarche de détermination du montant maximal manipulé

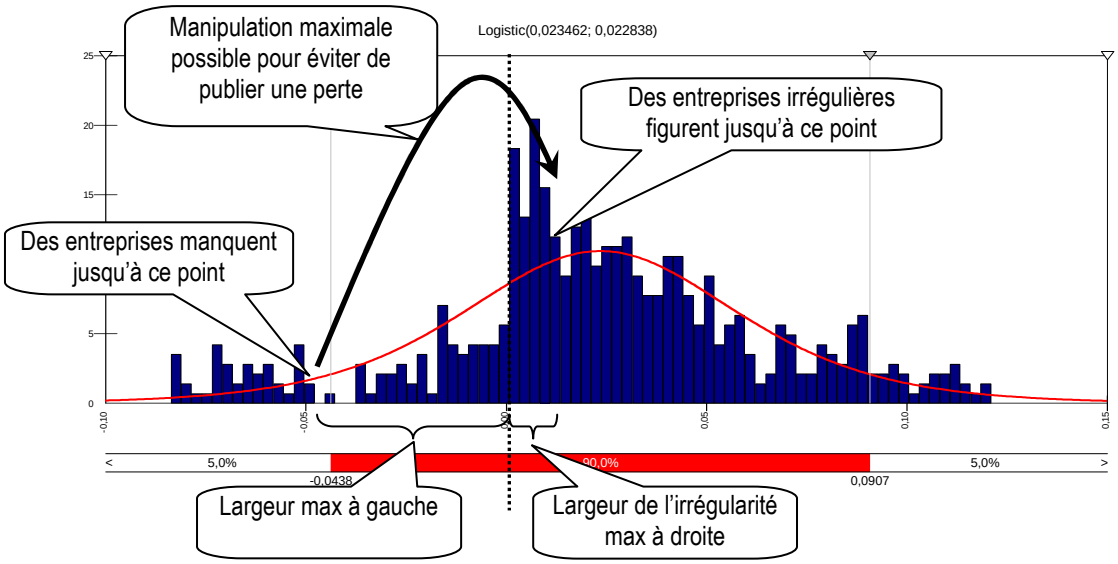


Tableau 102 : Nombre total d'entreprises bénéficiaires manipulatrices comparé au nombre maximum d'entreprises déficitaires manquantes, montant maximal manipulé

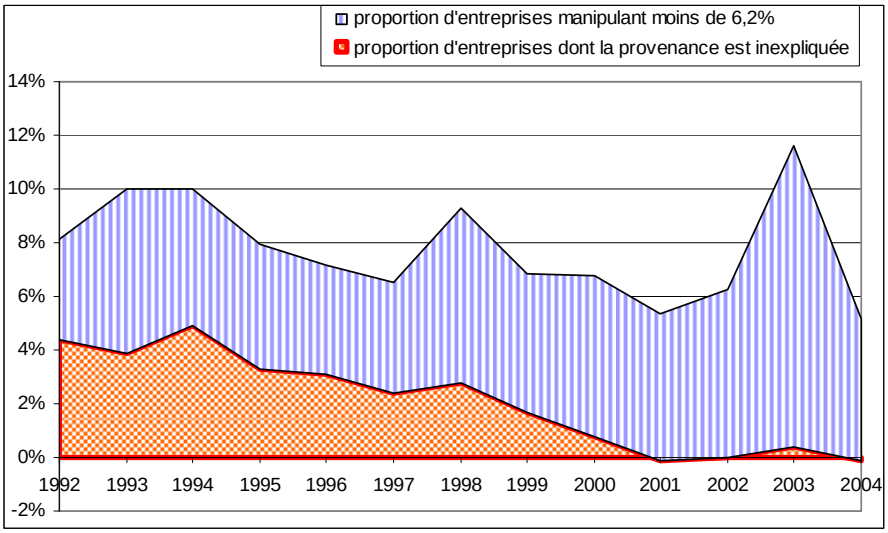
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Irrégularité à droite (surnombre max)													
Largeur	0,007	0,007	0,013	0,013	0,014	0,013	0,020	0,023	0,013	0,012	0,010	0,01845	0,017
Nombre d'entreprises en surnombre	**23.7**	**29.65**	**37.51**	**33.62**	**44.31**	**45.69**	**70**	**54.45**	**53.4**	**39.41**	**43.45**	**76.26**	**31.35**
Largeur max à gauche nécessaire pour compenser le surnombre à droite													
Largeur nécessaire (à gauche)	0,1	0,1	0,07	0,07	0,1	0,1	0,06	0,1	0,1	0,035	0,041	0,1	0,030
Sous-effectif maximal	**-10.92**	**-18.24**	**-19.05**	**-19.74**	**-25.08**	**-28.72**	**-49.27**	**-41.13**	**-47.07**	**-40.29**	**-43.67**	**-73.6**	**-32**
Le sous-effectif à gauche peut-il expliquer le surnombre à droite ?	NON	NON	NON	NON	NON	NON	NON	NON	Approx. oui	OUI	OUI	Approx. Oui	OUI
Écart inexpliqué	12.78	11.41	19.45	15.61	19.06	16.97	32.3	13.13	6.33	-0.88	-0.22	2.66	-0.65
% de l'irrégularité positive inexpliqué	53.9%	38.5%	49.2%	41.3%	43.4%	37.1%	29.6%	24.5%	11.9%	-2.2%	-0.5%	3.5%	-2.1%
! si >10% ; !! si >20% ; !!! si >30% ; !!!!si>40%	!!!!	!!!	!!!!	!!!!	!!!!	!!!	!!	!!	!				
Montant maximal manipulé (en % de l'AT)	>10,7%	>10,7%	8,3%	8,3%	>11,4%	>11,3%	8%	>12,3%	11,3%	4,7%	5,1%	11,8%	4,7%

En 1992 par exemple, près de 24 entreprises sont irrégulièrement faiblement bénéficiaires. La distribution des entreprises déficitaires montre que jusqu'à 11 entreprises (pour une largeur de 0,1) sont manquantes à gauche du seuil. Autrement dit, en 1992, il est impossible de compenser la totalité de l'irrégularité à droite par une irrégularité à gauche. Autre exemple : En 2001, près de 40 entreprises sont irrégulièrement faiblement bénéficiaires. En élargissant l'intervalle d'observation à gauche du seuil à 0,035, il apparaît un déficit de plus de 40 entreprises. Ainsi, en 2001, le surnombre d'entreprises irrégulièrement bénéficiaires peut être compensé par l'absence d'entreprises dont le résultat non manipulé serait une perte pouvant s'élever jusqu'à 3,5 % du montant du total actif.

L'intervalle d'observation n'est pas élargi au-delà de 0,1 à gauche du seuil car il apparaît qu'élargir l'intervalle au-delà ne permet généralement plus d'augmenter le nombre d'entreprises manquantes. Les années où la largeur maximale à gauche dépasse 0,1 l'hypothèse que les entreprises faiblement bénéficiaires sont uniquement des entreprises qui tentent de masquer un déficit ne peut pas être défendue. Raisonner en termes de montant manipulé devient alors délicat. Dans le Tableau 103, les montants sont précédés du signe « > ». Cela signifie que si des entreprises déficitaires manipulent leurs comptes

pour publier des bénéfices, les montants manipulés peuvent être supérieurs à 10 % du total actif. Mais l'étude ne peut déterminer ce montant.

Figure 131 : représentation graphique de la part des entreprises déficitaires identifiées comme manipulant leurs comptes pour éviter le seuil du résultat nul, par rapport au total des entreprises irrégulièrement faiblement bénéficiaires



Sur la plus grande partie de l'étude, les entreprises en sous-effectif à gauche sont toujours trop peu nombreuses pour expliquer le surnombre à droite. Autrement dit, le travail mené ne permet pas d'identifier l'origine de toutes les entreprises publiant un résultat irrégulièrement faiblement bénéficiaire. Ce constat est particulièrement vrai les toutes premières années de l'étude (de 1992 à 1997). Autrement dit, le principal résultat de ce travail est d'infirmer l'hypothèse de manipulation marginale pour éviter de publier une perte. Si les entreprises faiblement bénéficiaires sont irrégulièrement nombreuses, elles ne peuvent représenter uniquement des entreprises qui auraient publié une petite perte avant manipulation.

d. Montant moyen manipulé

Deux démarches peuvent être mises en œuvre pour déterminer le montant moyen manipulé chaque année par les entreprises qui évitent de publier une perte.

d.1. détermination approximative du montant moyen manipulé

La première estimation approximative consiste à considérer que la répartition des entreprises manquantes à gauche et des entreprises en surnombre à droite est répartie

de manière à peu près homogène. Autrement dit, le montant moyen est égal au montant maximal déterminé dans la sous-partie précédente divisé par deux (voir Figure 132).

Figure 132 : Illustration de la détermination du montant moyen manipulé sous hypothèse de répartition homogène des irrégularités

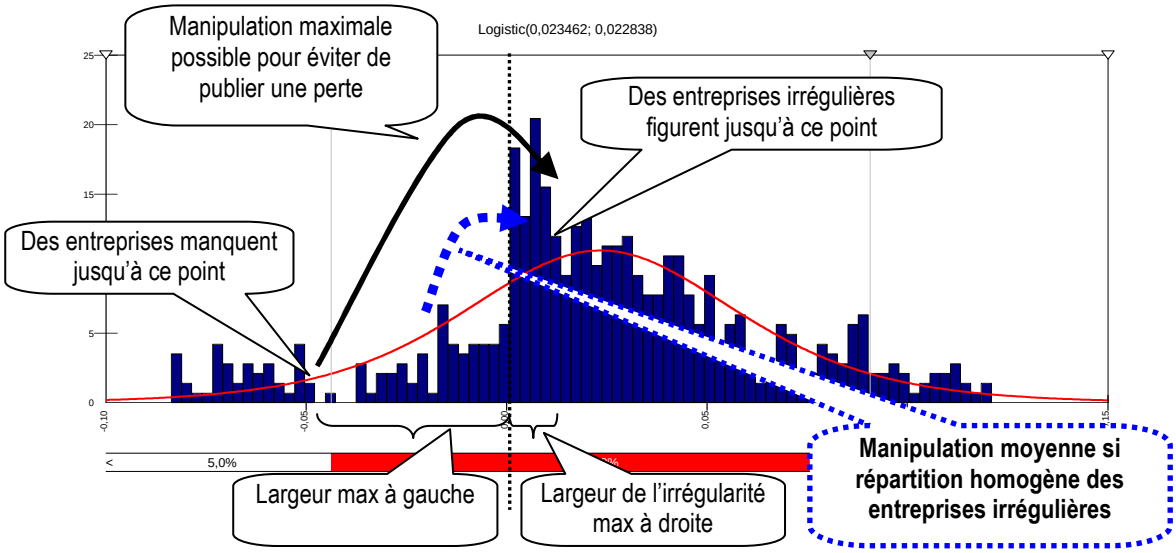


Tableau 103 : Calcul approximatif du montant moyen annuel manipulé par les entreprises évitant une perte

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
manipulation maximale (en % du total actif)	>10.7%	>10.7%	8.3%	8.3%	>11.4%	>11.3%	8.0%	>12.3%	11.3%	4.7%	5.1%	11.8%	4.7%
Manipulation moyenne (en % du total actif)	*5.4%*	*5.4%*	*4.2%*	*4.2%*	*5.7%*	*5.7%*	*4.0%*	*6.2%*	**5.7%**	**2.4%**	**2.6%**	**5.9%**	**2.4%**
Le sous-effectif à gauche peut-il expliquer le surnombre à droite ?	NON	NON	NON	NON	NON	NON	NON	NON	Approx. oui	OUI	OUI	Approx. Oui	OUI

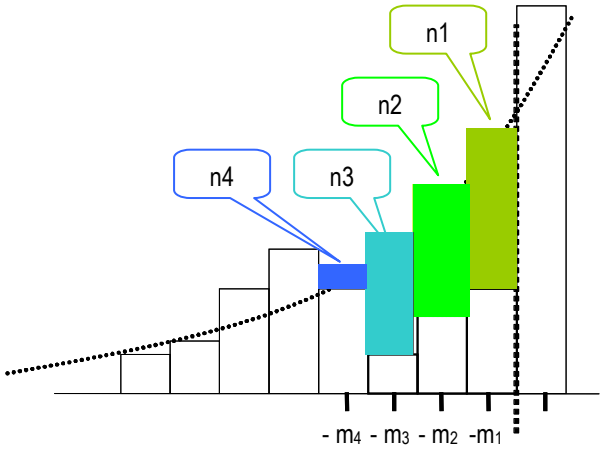
Les résultats des années 1992 à 1999 sont délicats à interpréter dans la mesure où le nombre d'entreprises déficitaires manquantes ne coïncide pas avec le surnombre d'entreprises bénéficiaires. Les trois années où le nombre d'entreprises déficitaires manquantes compense parfaitement le surnombre d'entreprises faiblement bénéficiaires (soit 2001, 2002 et 2004), le montant moyen manipulé semble porter sur près de 2,5 % du total actif.

d.2. détermination du montant moyen manipulé par moyenne pondérée

La seconde démarche de détermination du montant moyen manipulé consiste à découper la distribution autour du seuil en intervalles, et de déterminer pour chaque intervalle le nombre d'entreprises irrégulières. Chaque intervalle correspond alors à un montant manipulé. Le montant moyen se déduit en calculant la moyenne pondérée des montants irréguliers (voir Figure 133).

Figure 133 : Principe de détermination du montant moyen manipulé par moyenne pondérée

Sur le graphe ci-dessous, le montant moyen manipulé = (n1 x m1) + (n2 x m2) + (n3 x m3) + (n4 x m4)



Les calculs effectués sont relativement fastidieux puisqu'il faut relever, pour chaque distribution, et pour chaque intervalle, le nombre d'entreprises irrégulières (« n_i » sur la Figure 133) ainsi que le montant manipulé (« m_i »). Par ailleurs, le nombre d'entreprises déficitaires ne suffit pas à compenser le surnombre d'entreprises faiblement bénéficiaires de 1992 à 2000 et en 2003. Les montants manipulés ne peuvent donc pas être établis pour la totalité des entreprises ces années-là.

Les calculs par moyenne pondérée n'ont donc été mis en œuvre que les années où le nombre d'entreprises manquantes parmi les déficitaires permet de compenser le surnombre d'entreprises faiblement bénéficiaires, soit en 2001, 2002 et 2004. Les calculs sont décrits ci-dessous pour la seule année 2001. Le Tableau 187 détaillant les calculs des trois années se situe dans l'Annexe 15 : Calcul du montant moyen annuel manipulé par les entreprises qui évitent une perte ; calcul par moyenne pondérée.

Le calcul est effectué en deux temps car les calculs d'irrégularité droite et gauche ont été menés jusque là séparément. Le montant manipulé par une entreprise qui évite de publier une perte est égal au montant manipulé pour atteindre zéro, plus le montant manipulé

pour publier le faible bénéfice. Autrement dit, dans un premier temps (Tableau 104), le bénéfice excédentaire moyen (c'est-à-dire le bénéfice qui ne devrait pas avoir été publié) est mis en évidence. Puis il est additionné à la perte caché (c'est-à-dire la perte qui aurait dû être publiée) par ces mêmes entreprises (Tableau 105).

Tableau 104 : Bénéfices excédentaires (moyenne pondérée) des soc. faiblement bénéficiaires ; 2001

Année 2001													**total**	
Intervalle « i »	1	2	3	4	5	6	7	8	9	10	11	12		
De	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,01	0,011		
À	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,01	0,011	0,012		
Observé	5	8	11	9	9	6	9	13	10	7	7	7		
Ajusté	4,25	4,39	4,53	4,68	4,84	5	5,17	5,35	5,54	5,74	5,94	6,16		
Surnombre(« n_i »)	0,75	3,61	6,47	4,32	4,16	1	3,83	7,65	4,46	1,26	1,06	0,84	39,41	
montant manipulé(« m_i »)	0,0005	0,0015	0,0025	0,0035	0,0045	0,0055	0,0065	0,0075	0,0085	0,0095	0,0105	0,0115		
n_i x m_i =	0,00037	0,00541	0,01617	0,01512	0,01872	0,0055	0,02489	0,05737	0,03791	0,01197	0,01113	0,00966	0,214245	bénéfices irrégulièrement publiés
moyenne pondérée													**0,005436**	**bénéfice moyen irrégulièrement publié**

La largeur maximale permettant de détecter une irrégularité à droite du seuil en 2001 étant de 0,012, le calcul du montant moyen pondéré par les entreprises irrégulièrement bénéficiaires est effectué sur 12 intervalles de largeur 0,001. La moyenne pondérée obtenue est de 0,005436, soit 0,5436 % de l'actif total.

Tableau 105 : Pertes cachées (moyenne pondérée) des entreprises faiblement bénéficiaires en 2001

Année 2001												**total**	
Intervalle « i »[183]	1	2	3	4	5	6	7	8	9	10	11		
De	-0,001	-0,009	-0,01	-0,011	-0,019	-0,02	-0,021	-0,029	-0,03	-0,031	-0,035		
À	0	-0,001	-0,009	-0,01	-0,011	-0,019	-0,02	-0,021	-0,029	-0,03	-0,031		
Observé	2	16	4	0	12	1	1	13	0	3	2		
Ajusté	4,12	28,93	3,17	3,09	21,98	2,45	2,39	17,18	1,93	1,89	7,16		
Manquant **(« n_i »)**	-2,12	-12,93	0,83	-3,09	-9,98	-1,45	-1,39	-4,18	-1,93	1,11	-5,16	-40,29	
n_i corrigé	-2,12	-12,93	0,83	-3,09	-9,98	-1,45	-1,39	-4,18	-1,93	1,11	-4,28	-39,41[184]	
Perte masquée**(« m_i »)**	-0,0005	-0,005	-0,0095	-0,0105	-0,015	-0,0195	-0,0205	-0,025	-0,0295	-0,0305	-0,033		
n_i x m_i =	0,00106	0,06465	-0,00789	0,03244	0,1497	0,02827	0,02849	0,1045	0,05693	-0,03386	0,14124	0,56556	(déficit irrégulièrement évité)
moyenne pondérée												**0,014351**	(déficit moyen irrégulièrement évité)
total moyen												**0,019787**	**Soit 1,98% de l'actif total**

[183] Pour déterminer le montant de la perte masquée, le nombre d'intervalles étant plus élevé (35 intervalles de largeur 0,001), certains ont été réunis afin de diminuer le nombre de colonnes du tableau de calcul.

[184] Les calculs effectués au préalable montrent que l'irrégularité à gauche du seuil en 2001 est visible jusqu'à une largeur de 0,035. Or le nombre d'entreprises manquantes sur cet intervalle (40,29) est légèrement supérieur au nombre d'entreprises excédentaires à droite du seuil (39,41 déterminé dans le Tableau 104). Le nombre d'entreprises manquantes sur le dernier intervalle (]-0,35 ;-0,31]) est donc corrigé de manière à ce que le nombre d'entreprises manquantes à gauche soit rigoureusement égal au nombre d'entreprises excédentaires à droite du seuil.

La largeur maximale permettant de détecter une irrégularité à gauche du seuil en 2001 est de 0,035. Le calcul moyen pondéré de la perte masquée par les entreprises irrégulièrement bénéficiaires effectué sur cette largeur est de 0,014351, soit 1,4351 % de l'actif total. Additionné au bénéfice moyen irrégulièrement publié, le montant moyen des manipulations pour éviter des pertes s'élève en 2001 à 1,98 % de l'actif total. Les mêmes calculs conduisent à estimer la manipulation moyenne en 2002 et 2004 à 2,5 % et 2,1 % du total actif.

Les mesures par moyenne pondérée permettent donc de constater que les montants approximatifs calculés précédemment (paragraphe d.1.) sont légèrement surestimés ; 2,4 % au lieu de 2 % en 2001 ; 2,6 % au lieu de 2,5 % en 2002 ; 2,4 % au lieu de 2,1 % en 2004. Si le calcul pondéré tend à diminuer les montants estimés sous hypothèse d'homogénéité de la distribution des irrégularités, les ordres de grandeur restent cependant similaires et ne modifient pas la conclusion de l'étude.

d.3. Synthèse de l'étude des montant manipulés

L'étude des montants manipulés par les entreprises évitant de publier une perte aboutit aux conclusions suivantes :

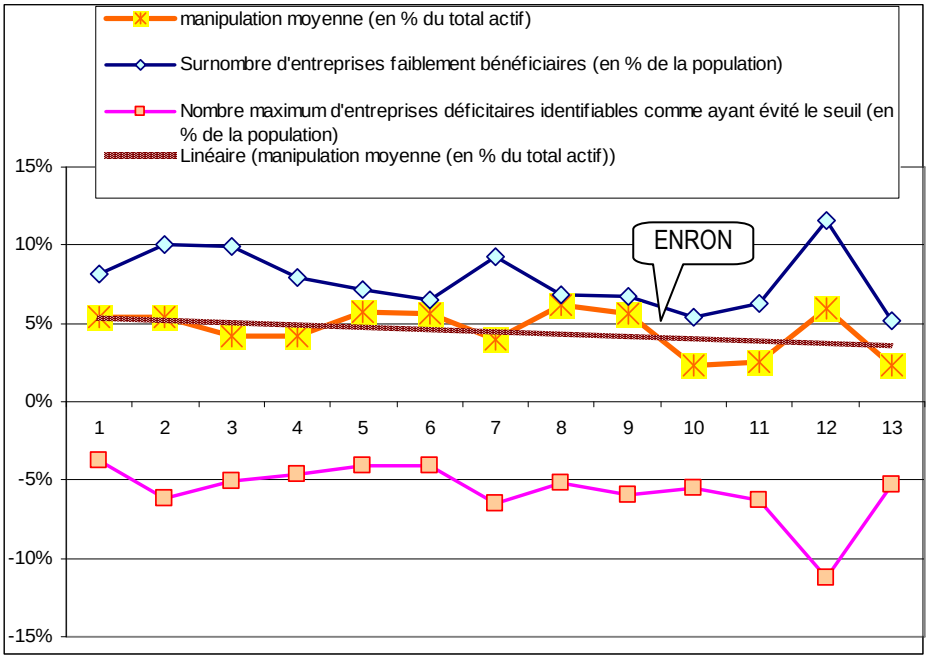
(1) Contrairement au postulat généralement admis dans la littérature comptable, les montants manipulés (pour éviter de publier une perte) ne peuvent pas être considérés comme marginaux. Les entreprises légèrement déficitaires manquantes dans les distributions ne sont jamais suffisamment nombreuses pour compenser le surnombre d'entreprises faiblement bénéficiaires. La « petite manipulation » effectuée à la marge par des entreprises qui ratent de peu un seuil, même si elle existe sans doute, ne peut expliquer à elle seule l'irrégularité constatée. Il apparaît au contraire que les entreprises déficitaires sont irrégulièrement peu nombreuses très loin du seuil. Il en découle que les montants manipulés pour éviter de publier une perte peuvent, pour certaines entreprises, concerner des montants très importants : plus de 10 % du total actif, soit plusieurs années de bénéfices moyens. Ces montants maximaux ne sont cependant que des cas extrêmes, et les montants moyens manipulés pour éviter une perte peuvent être estimés approximativement à 2,5 % du total actif[185]. Ces montants

[185] L'AT moyen étant de 7 455 millions d'euros (voir Tableau 11), les montants manipulés peuvent être estimés en moyenne à 186 millions d'euros. Soit de l'ordre de 8,5 milliards d'euros manipulés annuellement par l'ensemble des entreprises qui évitent de publier une perte (7,5% x 8 000 entreprises x 186 millions / 13 ans). Ce chiffre peut être comparé avec l'estimation de 30 milliards de dollars faite par Chen, Lin, Wang et Wu (2005) sur un échantillon d'entreprises américaines, mais avec une méthodologie différente.

sont extrêmement importants s'ils sont mis en relation avec la valeur moyenne (ou médiane) du résultat (voir Tableau 13) comprise selon les années entre 2 et 5 %. Bien entendu, ces manipulations ne concernent qu'une part réduite de la population (5 à 10 %). Mais cela signifie que les montants manipulés par ces quelques entreprises sont de l'ordre d'une année de résultat moyen. Autrement dit, les entreprises qui évitent de publier une perte n'hésitent pas à grever leurs résultats futurs d'un montant approximativement égal à une année de résultat.

(2) Les montants manipulés pour éviter de publier une perte tendent à diminuer au cours de la période étudiée, et tout particulièrement à partir de 2001. De 2001 à 2004, les manipulations pour éviter le seuil du résultat nul peuvent ne concerner exclusivement que des entreprises faiblement déficitaires (avant manipulations) alors que dans la décennie 1990, ces entreprises faiblement déficitaires ne peuvent en aucun cas suffire à expliquer l'irrégularité autour du seuil.

Figure 134 : Évolution des montants moyens manipulés (calcul approximatif) par les entreprises évitant de publier une perte



Autrement dit, il est intéressant de constater que l'année 2000 semble constituer un pivot. Après 2001, il est possible de penser que des entreprises continuent d'éviter de publier une perte, mais uniquement si la manipulation nécessaire est relativement marginale.

Après le scandale américain d'Enron, les entreprises françaises cotées semblent avoir gagné en « prudence ».

e. Modification du seuil de significativité

Comme cela a été développé (dans la sous partie c), le seuil minimum de 40 % retenu pour considérer que l'irrégularité est significative, est sévère, surtout en ce qui concerne l'irrégularité gauche. Comme l'irrégularité gauche est considérée comme non significative 6 années sur les 13 de l'étude (voir Tableau 101), le seuil de significativité a été diminué (de -40 % à -30 %) pour tester la robustesse des résultats. L'objectif est de vérifier dans quelle mesure ils sont sensibles au seuil de significativité retenu. Le Tableau 106 présente les irrégularités maximales mesurées avec ce nouveau seuil.

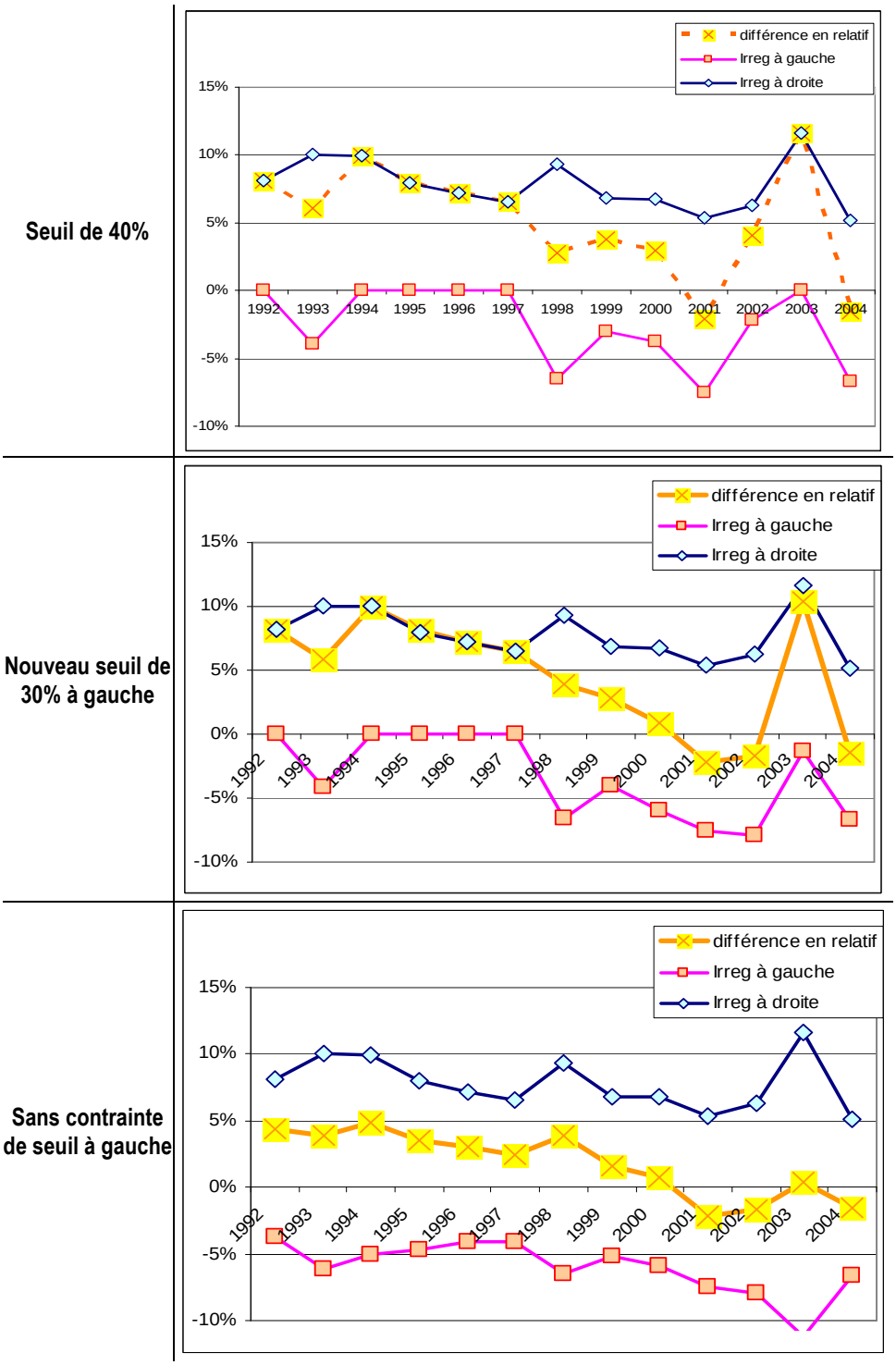
Tableau 106 : Largeur maximale permettant de constater une irrégularité et mesure des irrégularités annuelles maximales (seuil de significativité de 30 %)

Irrégularité à gauche	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Largeur correspondante	x	0.02	X	x	X	x	0.06	0.05	0.1	0.061	0.059	0.01	0.06
Nombre d'entreprises max	0	-12.36	0	0	0	0	-49.27	-32.04	-47.07	-55.12	-55.11	-8.46	-40.57
Nombre d'entreprises	291	296	376	422	616	702	753	797	791	734	693	657	609
Mesure relative	0.00%	-4,18%	0.00%	0.00%	0.00%	0.00%	-6,54%	-4,02%	-5,95%	-7.51%	-7,95%	-1,29%	-6.66%

Les irrégularités à gauche sont généralement moins prononcées que les irrégularités à droite du résultat nul. Pourtant, même de faible ampleur, elles sont toujours négatives. En abaissant le seuil de significativité de l'irrégularité gauche, le nombre d'années où les irrégularités gauche sont significatives augmente donc. Mais en diminuant le seuil de 40 à 30 %, ce nombre ne passe que de 7 à 8 (l'irrégularité de l'année 2003 est proche mais inférieure au seuil de 40 % ; elle est donc récupérée). En revanche, les irrégularités des années 1992 puis 1994 à 1997, même si elles sont toujours négatives (donc conformes à la théorie) apparaissent toujours trop faibles.

En abaissant le seuil, la largeur maximale de l'intervalle à gauche du résultat nul permettant de constater une irrégularité tend à augmenter, et certaines années (2000, 2001, 2003 et 2004), le nombre d'entreprises manquantes est donc légèrement réévalué à la hausse. La nouvelle évolution des irrégularités présentée sur la Figure 135 peut donc être comparée avec la Figure 128.

Figure 135 : Comparaison de l'évolution des irrégularités lorsque le seuil de significativité est modifié



Il apparaît que modifier le seuil de significativité de l'irrégularité à gauche ne modifie en aucun cas la tendance constatée : si le nombre d'entreprises évitant de publier une perte tend légèrement à diminuer sur la période, il apparaît que le nombre d'entreprises manquantes à gauche du seuil tend à augmenter au cours de la période. Autrement dit, les entreprises faiblement bénéficiaires ne peuvent toujours pas, malgré la modification du

seuil, être des entreprises évitant de publier des légères pertes avant l'année 2000. Quel que soit le seuil retenu, l'année 2003 apparaît comme étant le seul point réellement atypique dans cette tendance. Une analyse plus approfondie (Tableau 107) des mesures effectuées cette année-là conduit à relativiser cette valeur.

Tableau 107 : Réexamen des mesures d'irrégularité en faisant varier la largeur d'intervalle en 2003

De	-0,005	-0,01	-0,015	-0,02	-0,03	-0,04	-0,05	-0,06	-0,07	-0,1
À	0	0	0	0	0	0	0	0	0	0
Observé	9	20	31	34	47	61	70	76	87	107
Ajusté	14,63	28,46	41,54	53,9	76,62	96,93	115,08	131,29	145,78	180,6
écart	-5,63	-8,46	-10,54	-19,9	-29,62	-35,93	-45,08	-55,29	-58,78	-73,6
relatif	-38%	-30%	**-25%**	-37%	-39%	-37%	-39%	-42%	-40%	-41%

En 2003, l'irrégularité négative flirte avec le seuil des 40 %, sauf pour une largeur d'intervalle de l'ordre 0,015 où l'irrégularité tombe à 25 %. Autrement dit, cette rupture dans l'importance de l'irrégularité conduit à minimiser la mesure de l'irrégularité maximale (8,46 entreprises) alors que l'irrégularité est très persistante et que pour une largeur de 0,1 plus de 73 entreprises sont manquantes à gauche du seuil. Autrement dit, si le seuil de significativité est abaissé à 25 %, la différence entre l'irrégularités gauche et l'irrégularité droite devient quasiment nulle. L'année 2003 n'est alors plus une année atypique, et ce résultat renforce au contraire la tendance observée. Le « pic » observé en 2003 est donc lié essentiellement à la sévérité du seuil imposé dans le cadre de l'étude. La tendance à la baisse du nombre d'entreprises pratiquant de grandes manipulations pour éviter de publier une perte n'est pas remise en cause.

3. Synthèse de l'étude longitudinale

Les résultats de l'étude longitudinale découlent des mesures non paramétriques effectuées dans la partie 2 et des mesures paramétriques effectuées dans la partie 3. Le premier constat révèle que, quelle que soit la méthode utilisée, les tendances générales sont les mêmes.

- De 1992 à 2004, les distributions des résultats des entreprises françaises cotées présentent des irrégularités autour du résultat nul ; une irrégularité positive au dessus du seuil et une irrégularité négative en dessous. Ce constat est conforme avec les résultats des études antérieures ; les entreprises faiblement bénéficiaires

sont anormalement trop nombreuses alors que les entreprises faiblement déficitaires sont anormalement peu nombreuses.

- Durant la période, l'irrégularité à droite tend à diminuer légèrement, alors que l'irrégularité à gauche tend à augmenter (en valeur absolue). Autrement dit, il y a de plus en plus d'entreprises faiblement déficitaires manquantes, alors que le nombre d'entreprises faiblement bénéficiaires en surnombre tend à diminuer.

Si ces constats sont communs aux différentes méthodes de mesure des irrégularités, les mesures paramétriques présentent cependant des avantages qui permettent d'approfondir l'analyse.

(1) Elles reposent sur un fondement théorique que n'ont pas les méthodes non paramétriques.

(2) Le calcul de l'effectif théorique utilisé pour mesurer l'irrégularité n'est pas influencé par le choix de la largeur des intervalles.

(3) L'effectif de la population étudiée n'influence pas les mesures.

(4) Les méthodes paramétriques permettent de représenter graphiquement les irrégularités et de les comparer à toutes les autres irrégularités qui se manifestent sur l'ensemble de la distribution. L'intérêt de la représentation graphique peut être souligné : l'irrégularité n'est pas seulement mesurée en valeur absolue, mais en valeur relative (par rapport à l'ensemble des irrégularités de la distribution). Cette démarche pallie les faiblesses des tests statistiques de significativité dans la mesure où observer qu'une irrégularité est la plus importante sur l'ensemble de la distribution est plus parlant et tout aussi riche d'enseignement.

(5) Enfin, les méthodes paramétriques permettent de mettre en évidence de nouveaux indicateurs d'irrégularité : identification du nombre total d'entreprises manipulatrices et estimation des montants manipulés.

L'étude du nombre total d'entreprises manipulatrices conduit aux conclusions suivantes :

(a) Comme il est possible de déterminer le nombre d'entreprises en sur ou sous-effectif sur la largeur de l'irrégularité, il devient possible de mesurer l'irrégularité, non plus en l'exprimant par une mesure « locale », mais en l'exprimant en fonction de l'ensemble de la population. Ainsi, Il apparaît une relative stabilité de l'irrégularité à droite. Chaque année,

5 à 10 % de la population publie un résultat anormalement faiblement positif. Ce surnombre tend à diminuer au cours de la période étudiée.

(b) L'irrégularité à gauche, c'est-à-dire le nombre d'entreprises faiblement déficitaires manquantes, connaît une évolution plus délicate à analyser. Deux phases sont observables :

- Jusqu'en 1997 (à l'exception de 1993), le nombre d'entreprises à gauche manquantes est non significatif. Cela signifie que l'hypothèse selon laquelle les entreprises cherchent à éviter de publier un résultat faiblement négatif n'est pas vérifiée. Autrement dit, la manipulation comptable pour dépasser le seuil du résultat nul n'est pas limitée à des toilettages des comptes marginaux.

- À partir de 1998, le nombre d'entreprises faiblement déficitaires devient significativement anormalement faible. La présomption de petites manipulations pour éviter le seuil est alors une explication possible.

(c) Comme l'irrégularité à droite du seuil du résultat nul est mesurée indépendamment de l'irrégularité à gauche, il est possible de comparer le nombre d'entreprises en sur-effectif (à droite) à celui des entreprises manquantes (à gauche). Dans la littérature, les auteurs présument que les manipulations comptables pour éviter le seuil du résultat nul sont de faible ampleur. Les entreprises faiblement déficitaires manquantes seraient donc celles qui se retrouvent en surnombre parmi les entreprises faiblement bénéficiaires. Autrement dit, l'irrégularité à droite devrait être sensiblement la même (au signe près) que l'irrégularité à gauche. Or durant toute la période étudiée, à l'exception de 2001 et 2004, les irrégularités à gauche sont toujours inférieures (en valeur absolue) aux irrégularités à droite. Autrement dit, si les manipulations pour dépasser le seuil du résultat nul semblent concerner une part assez constante de la population, cette pratique ne concerne pas uniquement les entreprises faiblement déficitaires, et ce, même après 1997.

(d) Enfin, il apparaît que les irrégularités gauches sont généralement moins prononcées, mais généralement plus « persistantes », c'est-à-dire visibles sur un intervalle plus large que les irrégularités droites. Cette première analyse est approfondie grâce au calcul des montants maximaux et moyens manipulés pour éviter de publier une perte.

L'étude des montants manipulés aboutit aux conclusions suivantes :

(i) Le principal résultat de l'étude est de constater que contrairement au postulat généralement admis dans la littérature comptable, les montants manipulés (pour éviter de

publier une perte) ne peuvent pas être considérés comme marginaux. La « petite manipulation » effectuée à la marge par des entreprises qui ratent de peu un seuil, même si elle existe, ne peut expliquer à elle seule l'irrégularité constatée. Il apparaît au contraire que les montants manipulés pour éviter de publier une perte peuvent, pour certaines entreprises, concerner plus de 10 % du total actif, soit plusieurs années de bénéfices. Ces montants maximaux ne sont cependant que des cas extrêmes, et les montants moyens manipulés pour éviter une perte (même s'ils ne concernent qu'une part réduite de la population) peuvent être estimés approximativement à 2,5 % du total actif, soit de l'ordre d'une année de résultat. Ce résultat n'avait encore jamais été constaté dans les recherches antérieures.

(ii) Les montants manipulés pour éviter de publier une perte tendent à diminuer au cours de la période étudiée, et tout particulièrement à partir de 2001. De 2001 à 2004, les manipulations pour éviter le seuil du résultat nul peuvent ne concerner que des entreprises (relativement) faiblement déficitaires (avant manipulations) alors que dans la décennie 1990, ces entreprises faiblement déficitaires ne suffisent pas à expliquer l'irrégularité autour du seuil. Autrement dit, l'année 2001 constitue un pivot. Il est possible d'interpréter les résultats en disant que après 2001, des entreprises continuent d'éviter de publier une perte, mais uniquement si la manipulation nécessaire est (relativement) faible. Après le scandale américain d'Enron, les entreprises françaises cotées semblent avoir gagné en « prudence ».

C. Conclusion de la partie 3 : Les seuils comptables en France

L'étude de la littérature sur les seuils comptables menée dans la partie 1 de la thèse révèle plusieurs limites méthodologiques dans les études des irrégularités de distribution des résultats annuels, limites qui remettent en cause les interprétations faites jusqu'alors. La partie 2 de la thèse montre que si les études sont si sensibles aux choix méthodologiques, cela est essentiellement dû à la nature non paramétrique des démarches mises en place jusqu'alors pour étudier les seuils comptables. La troisième partie de la thèse propose une réponse à ces limites en développant une démarche paramétrique de mesure des irrégularités.

La section A du chapitre VII montre qu'il est impossible d'appliquer le théorème central limite. Par ailleurs, chaque année la distribution des résultats présente des caractéristiques récurrentes (allure d'entonnoir renversé, dissymétrie, concentration) qui ne peuvent être réduits à des manifestations de manipulations comptables. Pour ces raisons, la distribution des résultats agrégés ne peut être assimilée à une distribution Normale.

L'originalité de la démarche mise en oeuvre dans la thèse est d'être empirique. C'est l'observation des distributions des résultats publiés qui permet de déduire une allure caractéristique (loi de distribution théorique) des résultats annuels des entreprises françaises. La démarche de recherche se fonde donc sur les propositions suivantes :

1. Il est possible d'ajuster empiriquement une loi de distribution des résultats publiés.
2. Cette loi (l'allure de la distribution) est invariable (aux paramètres près). C'est une loi économique.

Pour être mise en oeuvre, la méthode doit faire face à deux problèmes : Comment agréger des distributions annuelles considérées comme des tirages indépendants ? Comment faire pour estimer une loi théorique des résultats non manipulés à partir d'observations de résultats publiés potentiellement manipulés ? Il s'avère que la « standardisation[186] » des distributions permet de répondre à ces deux préoccupations. En conclusion, l'ajustement mathématique de la loi de distribution théorique est réalisé sur

[186] Centrage des distributions annuelles sur la médiane annuelle, et réduction par l'écart type annuel.

la distribution agrégée des distributions annuelles centrées et réduites des résultats (préalablement mis à l'échelle) publiés par les entreprises.

Le chapitre VIII met en oeuvre la démarche et conclut que la distribution théorique des résultats inférieurs à la médiane (demi-distribution gauche) suit une loi Pareto2. Les résultats apparaissent d'autant plus robustes qu'ils sont valables aussi bien pour les entreprises françaises que britanniques et allemandes.

Une estimation moins précise peut être effectuée pour la seule partie centrale de la distribution en retenant la loi Logistique (plus pointue que la loi Normale, mais moins que la loi de Student). Ce résultat annexe mérite d'être souligné dans la mesure où l'étude de la partie centrale de la distribution peut sembler plus aisée dans le cadre de futures recherches. L'ajustement par la loi Pareto2 nécessite en effet le découpage de la distribution par rapport à la médiane, puis la transposition des données négatives (multiplication par « -1 »). Les variables extrêmes enfin peuvent influencer les calculs. Ce travail est lourd, et les graphes doivent être inversés avant d'être interprétés (puisque les données sont transposées). Tant que la zone étudiée (irrégularité au seuil du résultat nul) est inclue dans la partie centrale de la distribution, l'approximation par la loi Logistique peut être acceptée. Mais cette démarche interdit d'élargir l'intervalle d'étude au-delà des bornes fixées, ce qui rend difficile l'étude des montants manipulés.

L'intérêt de ces résultats mérite d'être mis en perspective avec les travaux menés dans la partie 2 de la thèse. En effet, aucun fondement théorique ne permettait de trancher en faveur d'une des différentes méthodes de mesure non paramétrique que sont la symétrie, l'interpolation ou l'extrapolation (linéaire, exponentielle ou logarithmique). L'ajustement global de la demi-distribution par la loi Pareto2 constitue en ce sens un progrès méthodologique dans l'estimation des irrégularités de distribution des résultats publiés.

Enfin, la méthode proposée est mise en application dans le cadre de l'étude longitudinale des résultats publiés par les entreprises françaises cotées de 1992 à 2004 dans le chapitre IX. Le premier constat révèle que, quelle que soit la méthode utilisée (paramétrique ou non paramétrique), les tendances générales sont les mêmes :

- De 1992 à 2004, les distributions des résultats des entreprises françaises cotées présentent des irrégularités autour du résultat nul ; une irrégularité positive au dessus du seuil et une irrégularité négative en dessous. Ce constat est conforme avec les résultats des études antérieures ; les entreprises faiblement bénéficiaires sont anormalement trop nombreuses alors que les entreprises faiblement déficitaires sont anormalement peu nombreuses.

- Durant la période, l'irrégularité à droite tend à diminuer légèrement, alors que l'irrégularité à gauche tend à augmenter (en valeur absolue). Autrement dit, il y a de plus en plus d'entreprises faiblement déficitaires manquantes, alors que le nombre d'entreprises faiblement bénéficiaires en surnombre tend à diminuer.

Si ces constats sont communs aux différentes méthodes de mesure des irrégularités, les mesures paramétriques permettent d'approfondir l'analyse en calculant de nouveaux indicateurs d'irrégularité : identification du nombre total d'entreprises manipulatrices et estimation des montants manipulés.

- Il apparaît ainsi que chaque année, 5 à 10 % de la population publie anormalement un petit bénéfice. Ce surnombre tend à diminuer au cours de la période étudiée. L'irrégularité à gauche, c'est-à-dire le nombre d'entreprises faiblement déficitaires manquantes, connaît en revanche une évolution plus délicate à analyser. Jusqu'en 1997 (à l'exception de 1993), ce nombre est peu significatif. Cela laisse supposer que peu d'entreprises cherchent à éviter de publier un résultat faiblement négatif. Autrement dit, la manipulation comptable pour dépasser le seuil du résultat nul n'est pas limitée à des toilettages des comptes marginaux. À partir de 1998, le nombre d'entreprises faiblement déficitaires manquantes devient plus significatif. La présomption de petites manipulations pour éviter le seuil est alors une explication possible.

- La comparaison du nombre d'entreprises en sur-effectif (à droite) à celui des entreprises manquantes (à gauche) montre que avant 2001, les irrégularités à gauche sont toujours inférieures (en valeur absolue) aux irrégularités à droite. Autrement dit, si les manipulations pour dépasser le seuil du résultat nul semblent concerner une part assez constante de la population, cette pratique ne concerne pas uniquement les entreprises faiblement déficitaires, et ce, même après 1997.

- Enfin, il apparaît que les irrégularités gauches sont généralement moins prononcées, mais généralement plus « persistantes », c'est-à-dire visibles sur un intervalle plus large que les irrégularités droites. Cela laisse penser que les entreprises irrégulières qui publient un petit bénéfice peuvent être des entreprises qui auraient publié une perte importante sans manipulation.

Cette première analyse est approfondie grâce au calcul des montants maximaux et moyens manipulés pour éviter de publier une perte. Ces calculs aboutissent aux conclusions suivantes :

- Le principal résultat de l'étude est de constater que contrairement au postulat généralement admis dans la littérature comptable, les montants manipulés (pour éviter de publier une perte) ne peuvent pas être considérés comme marginaux. La « petite manipulation » effectuée à la marge par des entreprises qui ratent de peu un seuil, même si elle existe, ne peut expliquer à elle seule l'irrégularité constatée. Il apparaît au contraire que les montants manipulés pour éviter de publier une perte peuvent, pour certaines entreprises, concerner des montants très importants : plus de 10 % du total actif, soit plusieurs années de bénéfices. Ces montants maximaux ne sont cependant que des cas extrêmes, et les montants moyens manipulés pour éviter une perte (même s'ils ne concernent qu'une part réduite de la population) peuvent être estimés approximativement à 2,5 % du total actif, soit de l'ordre d'une année de résultat. Chaque année, près de 8,5 milliards d'euros seraient ainsi manipulés à la hausse par les entreprises cotées françaises qui évitent de publier une perte.

- Les montants manipulés pour éviter de publier une perte tendent à diminuer au cours de la période étudiée, et tout particulièrement à partir de 2001. Les manipulations pour éviter le seuil du résultat nul peuvent alors ne concerner que des entreprises (relativement) faiblement déficitaires (avant manipulations). Autrement dit, l'année 2001 constitue un pivot. Après 2001, des entreprises continuent d'éviter de publier une perte, mais uniquement si la manipulation nécessaire est relativement marginale. Après le scandale américain fortement médiatisé d'Enron, les entreprises françaises cotées semblent avoir gagné en « prudence ».

CONCLUSION GÉNÉRALE

La recherche menée dans le cadre de la thèse s'intègre dans le cadre général de la littérature académique sur la qualité des informations comptables, cadre qui résonne avec l'actualité depuis les scandales financiers de ces dernières années d'une part, et l'accélération du processus d'harmonisation internationale des normes comptables d'autre part. Ainsi, si les irrégularités de distribution sont des indicateurs de mauvaise qualité de l'information comptable, leur étude doit permettre d'améliorer la connaissance sur le thème plus général de la qualité de l'information comptable.

Depuis plusieurs années, il est possible de constater un intérêt croissant pour l'étude des seuils, et plus spécifiquement dans les publications internationales. Cette littérature a pu parfois sembler prendre le relais des modèles d'identification des pratiques de gestion du résultat via les calculs d'accruals. La thèse qui est présentée ici tente d'apporter un éclairage original sur ce sujet.

Les seuils comptables, et la présomption de manipulation qui leur est associée, sont généralement présentés comme un cas particulier des pratiques de gestion du résultat. À titre d'exemple, Healy et Whalen (1999) dans leur revue de littérature leur consacrent moins d'une page après avoir consacré une douzaine de pages à décrire les motivations à la gestion du résultat. Quant aux montants manipulés, réputés marginaux, ils semblent reléguer l'étude des seuils au niveau de celui de curiosité. La thèse présentée espère contribuer à modifier ce regard.

De l'importance des seuils comptables

Les entreprises qui manipulent leurs comptes peuvent le faire à la hausse ou à la baisse. La motivation pour manipuler ses comptes à la baisse est de dégager des marges de manœuvre en période faste (ou à l'occasion d'un événement exceptionnel) pour faire face à d'éventuelles contre-performances futures. À la hausse, l'objectif est de masquer une contre-performance ponctuelle en utilisant les « réserves » passées ou en espérant que

les résultats futurs permettront de compenser cette manipulation. Mais que ce soit à la hausse ou à la baisse, il est possible de présenter le problème autrement : Quel montant manipuler ?

Deux réponses sont théoriquement envisageables : L'entreprise peut, soit manipuler le montant le plus important possible (à la baisse ou à la hausse), soit chercher à manipuler un montant jugé raisonnable, suffisant pour atteindre un objectif qu'elle se fixe.

Le premier cas de figure (manipulation maximale dans la limite du possible) est théorique. En pratique, on imagine mal l'intérêt pour une entreprise de manipuler le résultat d'un montant infini si ce n'est dans des situations très particulières. En effet, une telle manipulation, non seulement est très risquée (elle a d'autant plus de chances d'être découverte qu'elle est importante, et ses sanctions en seront d'autant plus graves). Mais par définition, de telles manipulations sont des opérations exceptionnelles, impossibles à reproduire l'année suivante et difficiles à masquer sur le long terme. Elles ne peuvent dès lors qu'être exceptionnelles. Il arrive que ponctuellement, une entreprise puisse être amenée à solder des risques passés et dégager le plus de marges de manœuvre possible pour le futur (sous contrainte de ne pas déposer le bilan pour autant). Les grands bains comptables, notamment lorsque l'entreprise traverse une crise, fait l'objet d'une restructuration ou d'un changement de dirigeant, font partie de ces cas, rares, qui émaillent ponctuellement la vie des entreprises. Hormis ces cas, manipuler à la hausse ou à la baisse d'un montant « le plus important possible » est difficilement justifiable.

Il en découle logiquement que presque toutes les manipulations de résultat doivent répondre à un objectif concret, atteignable par un montant défini (sans présumer de l'importance de ce montant). Autrement dit, dans la plus grande majorité des cas, la gestion du résultat est réalisée pour atteindre un objectif.

Faut-il alors faire une distinction entre le terme d'objectif, et le terme de seuil ? Dans la littérature, le terme de seuil comptable n'a été jusqu'alors employé que pour désigner trois objectifs particuliers : le résultat nul, la variation nulle de résultat et l'erreur nulle de prévision. Le terme d'objectif lui est d'ailleurs souvent indifféremment substitué. Mais si la notion de seuil se confond avec celle d'objectif, et que les seuils comptables ne se limitent pas aux trois seuls seuils suscités, alors la place de l'étude des seuils comptables doit être réévaluée.

La thèse vise à apporter des éléments nouveaux à cette réflexion conceptuelle. Ces éléments sont tout autant d'ordre sémantique et méthodologique qu'empirique.

Apport sémantique

L'apport de la thèse est tout d'abord sémantique. La partie I propose une clarification terminologique, dresse une typologie des seuils, et élabore un cadre d'analyse.

La clarification terminologique précise les définitions des concepts utilisés dans la littérature sur les seuils. Si les discontinuités ne sont pas les seules irrégularités, elles signalent par leur récurrence des seuils de publication. La notion d'effet de seuil permet de traduire la relation non linéaire entre coût du financement et résultat publié. Mais les motivations financières ne sont pas les seules, et des motivations contractuelles ou réglementaires peuvent intervenir dans les décisions d'évitement des seuils comptables. Enfin, les facteurs psychologiques qui pèsent sur les investisseurs sont distingués de ceux qui pèsent sur les dirigeants.

La typologie des seuils montre que les trois seuils étudiés jusqu'alors dans la littérature ont en commun la caractéristique de pouvoir être mis en évidence par des méthodes statistiques à partir de bases de données. Ce sont donc des objectifs repérables grâce aux irrégularités de distributions. Mais la typologie conduit à prédire l'existence de certaines pratiques de gestion du résultat pour éviter des seuils relatifs (notamment d'origine contractuelle ou réglementaire) encore jamais observés empiriquement.

Le cadre théorique d'analyse des seuils de publication comptable dressé dans la partie 1 permet, enfin, d'articuler les explications psychologiques avec le cadre d'analyse plus général de la recherche sur la gestion du résultat. Le cadre de la théorie politico-contractuelle (théorie positive), habituellement mobilisé dans les recherches comptables sur base de données, ne permet pas, en effet, d'expliquer à lui seul l'existence des seuils. Il doit être complété par des théories issues de la psychologie, et par l'étude de la formation du jugement des dirigeants. Le cadre proposé met en relief le rôle clef des anticipations des dirigeants, et permet de mieux appréhender certains résultats secondaires figurant dans des études publiées, mais non interprétés. Une relecture de ces résultats peut être alors envisagée. Il est notamment intéressant de comparer les surcoûts d'origine financière, réglementaire ou contractuelle selon la taille, le secteur ou l'endettement des entreprises. De même, l'influence des facteurs psychologiques dans la formation du jugement des dirigeants mérite d'être étudié. Le cadre théorique proposé permet ainsi d'envisager de nouvelles perspectives d'études sur ces sujets.

En conclusion, la réflexion sémantique permet d'élargir le champ généralement affecté aux études sur les seuils.

Apport méthodologique

S'il est simple de constater graphiquement les irrégularités de distribution, effectuer des comparaisons, ou identifier des déterminants, nécessite la mise en place d'indicateurs de mesure de l'amplitude des irrégularités fiables et comparables. Or les mesures non paramétriques des irrégularités utilisées jusqu'alors dans la littérature comptable manifestent plusieurs limites. La partie 2 de la thèse précise ces limites, et identifie leur impact concret à partir de l'étude des résultats publiés par les entreprises cotées françaises de 1992 à 2004.

L'impact de trois choix méthodologiques sont ainsi étudiés : le choix d'un instrument de mesure (symétrie, interpolation, extrapolation), le choix de la largeur d'intervalle retenue pour effectuer les mesures, et le choix de l'indicateur de taille retenu pour mettre le résultat à l'échelle. Il apparaît ainsi que les mesures non paramétriques effectuées, même si elles révèlent toujours des irrégularités de même signe, sont très sensibles à ces choix. Autrement dit, le choix de la méthode d'étude, sans remettre en cause l'existence des irrégularités de distribution des résultats publiés, influence très sensiblement les conclusions des recherches effectuées sur le sujet.

Afin de répondre à ces limites, la thèse propose dans la partie 3 une nouvelle démarche de mesure des irrégularités. En effet, il apparaît que les limites des mesures effectuées jusqu'alors dans la littérature sont essentiellement liées à leur nature non paramétriques ; les estimations théoriques des effectifs sont estimées à partir d'interpolations locales. La démarche proposée est doublement innovante : elle est paramétrique (elle se fonde sur la connaissance de la loi de distribution attendue des résultats publiés) et elle est empirique (aucune hypothèse a priori n'est formulée sur la loi de distribution qui découle de l'ajustement des distributions observées après standardisation). Concrètement, la demi distribution gauche est ajustée par une loi Pareto2, et les mesures ainsi effectuées, tout en étant corrélées aux mesures non paramétriques, révèlent des évolutions moins volatiles. Cette allure de distribution est constatée aussi bien sur des distributions de résultats d'entreprises françaises cotées, que britanniques ou allemandes.

Enfin, au-delà de la comparaison avec les méthodes non paramétriques, les mesures paramétriques permettent enfin de faire apparaître deux nouveaux indicateurs utiles à la compréhension des seuils comptables. En s'affranchissant des contraintes de largeur d'intervalle, elles permettent de maximiser les mesures d'irrégularités, et d'en déduire le nombre total d'entreprises manipulatrices et les montants manipulés.

Si les limites méthodologiques ont parfois été utilisées par des Cassandres pour remettre en cause l'intérêt de l'étude des seuils comptables, la thèse prend le contre-pied de cette attitude et tente de dépasser ces faiblesses pour rebondir et faire avancer la recherche sur le sujet. Les propositions méthodologiques permettent ainsi d'élargir le champ des études sur les seuils.

Apport empirique

Si l'apport méthodologique de la thèse peut sembler le plus important, et occupe la majorité du document final, il ne doit pas faire oublier que l'apport de la thèse est également empirique. En appliquant la nouvelle méthodologie proposée à l'étude des manipulations comptables pour éviter de publier des pertes par les entreprises françaises cotées de 1992 à 2004, la thèse réalise une étude longitudinale jamais réalisée auparavant en France.

Concrètement, il apparaît ainsi que 5 à 10 % des entreprises cotées françaises manipulent chaque année leur résultat pour éviter de publier une perte. Ce nombre tend à diminuer durant la période d'étude. Par ailleurs l'étude montre que, avant 2001, le nombre d'entreprises irrégulièrement bénéficiaires n'est jamais compensé par le sous-effectif d'entreprises déficitaires. Autrement dit, avant 2001, la présomption de manipulation marginale ne peut être vérifiée. Si des entreprises manipulent pour publier un petit bénéfice, il n'y a aucune raison de croire que ces entreprises soient essentiellement des entreprises qui auraient publié une petite perte avant manipulation.

Après 2001, ce constat s'atténue, et le nombre d'entreprises publiant irrégulièrement un petit bénéfice peut être compensé par les sous-effectifs constatés parmi les entreprises déficitaires. Autrement dit, après 2001, l'interprétation d'une manipulation relativement faible pour éviter de publier une perte ne peut pas être rejetée.

L'étude des montants manipulés confirme et étend ce résultat. Si les entreprises irrégulièrement faiblement bénéficiaires manipulent leur résultat à la hausse, alors les montants moyens manipulés tendent à diminuer au cours de la période d'observation, notamment après 2001, pour s'élever à un ordre de grandeur proche d'une année de résultat. Ces montants moyens ne doivent cependant pas masquer des comportements très hétérogènes, et l'étude révèle que les manipulations effectuées à la hausse par certaines entreprises peuvent s'élever à plusieurs années de résultats. Enfin, le montant

total annuel de ces manipulations peut être évalué de l'ordre de 8,5 milliards d'euros chaque année, pour les seules entreprises françaises cotées.

Au-delà de ces éléments factuels, le principal résultat de cette étude dépasse le cadre de la France. En effet la thèse, en remettant en cause le postulat de manipulation marginale, souligne l'importance de la recherche sur les seuils. La gestion du résultat pour dépasser des seuils n'apparaît pas comme un cas spécifique et marginal de gestion du résultat, mais comme une pratique importante, voire majoritaire. Cela va dans le sens du constat réalisé dans l'enquête de Graham, Harvey et Rajgopal (2005) réalisée directement auprès des dirigeants. Ils contribuent également à expliquer que des différences entre les performances futures des entreprises légèrement bénéficiaires d'une part, et légèrement déficitaires d'autre part puissent être identifiées dans les études de Degeorge, Patel et Zeckhauser (1999) et de Mard (2004).

En conclusion…

En conclusion, les apports sémantique, méthodologique et empirique de la thèse contribuent à élargir le regard porté sur les seuils comptables. Ils intéressent en premier lieu les chercheurs en comptabilité qui peuvent utiliser ou enrichir le cadre explicatif proposé, et adopter ou améliorer la démarche paramétrique développée. D'un point de vue managérial, les investisseurs et les analystes financiers qui tentent d'interpréter au mieux les informations financières s'intéressent de près aux conséquences, pour une entreprise, d'être confrontée à un seuil. L'étude de l'évolution des irrégularités en France, et l'importance relative des montants potentiellement manipulés contribuent à mieux interpréter les résultats publiés et en déduire des décisions plus pertinentes en matière de choix d'investissement, de cotation boursière ou de prévision des performances futures. Ce travail intéresse également les normalisateurs comptables qui cherchent à améliorer la qualité des informations financières communiquées par les entreprises. Enfin, si les effets d'évitement sont des irrégularités liées à des appréhensions naturelles, de nature symbolique et/ou psychologique, les identifier, les analyser, et communiquer sur leurs causes peut permettre de démystifier ces seuils, aussi bien auprès des dirigeants que des investisseurs, et par voie de conséquence, de participer à leur atténuation. Il faut souhaiter que la thèse contribue à ce processus.

ANNEXES

Annexe 1 : Correction de la base[187]

Un certain nombre de tests ont été menés dans le but de corriger des informations manifestement erronées présentes dans la base. Ces corrections concernent essentiellement la variable résultat net des entreprises françaises qui est au cœur des préoccupations.

1. Résultat strictement égal à zéro

Quatre entreprises publient au moins une fois un résultat strictement égal à zéro durant la période.

Tableau 108 : Entreprises publiant un résultat strictement nul

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
CAISSE NATIONALE DES AUTOROU		0	0	0	0	0	0	0	0	0	0
MARSEILLAISE DU TUNNEL PRADO (smtpc)	0	0	0	0	0	0	0	0	0	6.69	6.47
RIGHINI SPA			0.86	0.64	0.52	0.53	0.64	0.44	0		
SOFIBUS SA			0	-0.01	-6.73	1.24	1.29	1.63	1.96	1.93	2.95

La Caisse Nationale des Autoroutes ne publie que des résultats strictement nuls. Créée en 1963, la CNA est un établissement public à caractère administratif doté de l'autonomie financière qui a pour mission de procurer aux sociétés concessionnaires d'autoroutes les ressources destinées au financement de la construction ou de l'aménagement des autoroutes donnant lieu à la perception de péages. Pour ce faire, elle émet des emprunts sur les marchés français et internationaux et contracte des emprunts auprès de certains établissements financiers. Du fait du statut particulier de cette société, il a été décidé de l'éliminer de l'étude.

La Société Marseillaise du Tunnel Prado Carénage est une société anonyme entièrement privée, concessionnaire de la Ville de Marseille pour une durée de 32 ans (jusqu'en 2025). Le 1° janvier 2001, la Communauté Urbaine de Marseille Provence Métropole est devenue autorité concédante du tunnel Prado Carénage à la place de la Ville de

[187] Détails des corrections abordées dans le chapitre V page 129.

Marseille. La société devient cotée en bourse à partir de Juillet 2004. Suite à la transformation de son statut, la société a publié des résultats non nuls. Il a été décidé de retirer de la base les résultats antérieurs à 2003 de cette société.

Dans la base Thomson Worldscope, la société Righini affiche au 31/12/2002 un résultat négatif de -0,00036 avant de quitter la base en 2003, alors que le résultat figurant dans la base Global Vantage est strictement nul. Cette très légère différence pose d'autant plus problème que l'observation se situe exactement sur le principal seuil d'étude. Si l'on retient le résultat strictement nul de la base Global, alors la société évite la publication du résultat négatif. Dans le cas contraire, elle le rate. Il a été impossible de vérifier directement le rapport annuel 2002 de la société Righini. La base est corrigée en inscrivant le résultat (très faiblement) négatif partant de l'hypothèse que si erreur de saisie il y a, elle est plus probablement faite en faveur de l'arrondis.

Le résultat strictement nul publié par la société Sofibus est confirmé en 1996. Les résultats de 1992 à 1995 (qui sont absents de la base Global Vantage, mais qui sont fournis par la base Thomson Worldscope) ne sont pas nuls. La valeur strictement nulle en 1996 a été laissée et les années 1992 à 1995 ont été complétées (pour mémoire, de 92 à 95, AT = 83,24 ; 80,88 ; 78,69 ; 81,45 et IB = 0,91 ; 0,97 ; 1,04 ; 0,77).

Cette première série de corrections conduit à diminuer la taille de l'échantillon à 1010 entreprises (au lieu de 1011), à supprimer 9 données, à en corriger 1, et à compléter 8 données manquantes (dont 4 résultats, et 4 total actif).

2. Actif total strictement égal à zéro

Une (et une seule) société présentait un total actif strictement nul dans la base de donnée pour la seule année 2000. Il s'agit de la société DECAN Groupe SA, spécialisée dans les technologies de l'information et de la communication. Cette donnée (78.70379) a été retrouvée dans la base Thomson Worldscope. La base est corrigée de cette valeur.

3. Données récentes

Les premières données enregistrées dans la base Global Vantage, et accessibles pour la recherche, datent de 1992. Lors de la collecte, en mars 2006, les données concernant l'année 2005 étaient encore extrêmement rares. La base est donc finalement constituée des données de 1992 à 2004. L'horizon d'analyse est donc de 13 ans.

4. Données manquantes

Un travail a été mené afin de compléter les données manquantes de la base dans la mesure du possible. Ce travail s'est avéré titanesque et chronophage. Il a cependant permis de récupérer près de 200 observations (soit près de 2,5 % de la population).

Il est inconcevable de vérifier exhaustivement les données de la base. Par ailleurs, en ayant retenu les entreprises actives et inactives, l'absence de donnée une certaine année ne signifie pas nécessairement que la donnée est manquante. L'entreprise peut ne pas encore exister, ou ne pas être encore cotée, ou avoir quitté la cote, ou avoir fait l'objet d'une fusion, d'une acquisition, ou d'une liquidation. Cela concerne près de 40 % des observations.

Les « vraies » données manquantes ont été identifiées de deux manières : (1) en identifiant les « trous » dans les séries de résultats, ou (2) en vérifiant la présence simultanée de la variable résultat avec une variable de mise à l'échelle.

a. Trous dans les résultats

La base Global Vantage présente des « trous » qui ne sont pas toujours explicables. Les cas étranges ont été identifiés par un test sous Excel consistant à identifier les absences de résultat une année lorsque le résultat était présent dans la base l'année d'avant, et l'année d'après. Le Tableau 109 présente les sociétés concernées.

Les données ont été ensuite complétées, société par société, en recherchant les informations directement dans les rapports financiers (téléchargés sur Internet) ou dans la base Thomson Worldscope. Ce travail de collecte est long, fastidieux, et pas nécessairement très productif.

Tableau 109 : Données manquantes identifiées comme des « trous » dans les séries de données

Dans le tableau, les données en gras sont les données complétées (base Thomson Worldscope ou rapports financiers disponibles sur Internet), les chiffres non gras sont les données de la base Global, les « ? » signifient que la variable n'a pas été modifiée (parce qu'elle n'a pas été trouvée, ou parce que la donnée trouvée n'est pas pertinente), les « X » signifient que la donnée a été complétée ou modifiée.

Company Name	IB1992	IB1993	IB1994	IB1995	IB1996	IB1997	IB1998	IB1999	IB2000	IB2001	IB2002	IB2003	IB2004	Commentaire
AB SOFT SA							X -1.087	X 0.063	-1.6	X -2.473	1.7			
AFIBEL								3.7	4.5	X 8.220	5.3	8.8	7.7	
AVANQUEST SOFTWARE SA								1.5	2.4	X -4.385	5.4	1.0	3.8	(ex bvrp)
AVENIR FINANCE									1.6	X 0.973	1.1	0.4	1.5	

AVENIR NUMERIC'S									0.41	X 0.703	1.15	1.60	1.72	
BUSINESS ET DECISION									1.5	X 2.068	-4.5	1.0	1.0	
CCF-CREDIT COMMERCIAL FRANCE	137.5	159.3	180.4	187.0	208.3	242.9	319.0	449.3	477.0	X 517.037	562.0	627.0	729.0	
CFCAL-ALSACE LORRAINE BANQUE									3.8	X 4.425	4.4	5.4	34.5	
CIE FRANCAISE MOULAGES METAL									1.1	X -5.205	-9.9	0.1	0.5	
CITEL SA									0.6	X -0.058	0.3	-1.2	-0.1	
CONSORT NT									1.22	X 1.744	1.59	1.01	0.89	
CRCAM ALPES-PROVENCE									45.93	X 50.452	62.18	69.92	78.68	
CRCAM NORD									41.56	X 100.737	102.41	126.69	140.94	
CROSS SYSTEMS CO								1.2	-1.5	X -5.365	-2.0	-5.4	-2.5	
CYBER PRESS PUBLISHING									-5.1	X -5.028	-12.8	4.4	-5.5	
ELIT GROUP							0.9	1.5	1.5	X 0.693	-9.3	X -2.792	-2.8 -1.710	clôture juin (Décal.)
EPC GROUPE			3.4	4.5	4.6	6.1	2.7	6.8	10.9	X 11.918	9.9	1.2	X 8.563	(explosifs)
ETUDES REALISATNS MOULES OUT							1.6	1.3	1.3	X 1.077	0.5	0.6	X 0.507	(ermo)
EUROCRYSTAL							X -0.01456	X 0.0217 7	0.01	X 0.06200	0.08	-0.03	-0.17	
EUTELSAT COMMUNICATION			50.2	65.5	78.9	128.4	133.2	155.9	316.3	?	226.5	59.6	269.8	
FONCIERE MASSENA	2.5	2.9	3.3	3.0	5.5	5.6	1.9	1.5	-0.4	X -0.364	13.3	16.0		
GAMELOFT									-2.4	-66.2	X - 23.241	2.4	1.9	
GAZ DE FRANCE			203.1	291.6	388.6	277.0	482.0	418.8	318.0	X 891.000	838.0	910.0	1 046.0	
GRAND LIVRE DU MOIS					3.7	2.6	0.7	-0.4	-4.4	X 1.045	3.2	1.5	X 3.064	(glm)
GROUPE CENTRE RECHERCHES IND							5.0	8.5	10.8	X 9.414	6.1	3.5	X 7.186	(crit)
GROUPE JAJ				3.7	? date	2.5	2.8	-0.4	-1.0	-0.2	1.2	1.1	X 1.063	Différences avec Thomson !
GUILLEMOT CORP SA							2.0	3.9	7.0	2.4	-49.6	X -32.378	-2.7	
IMMEUBLES DE FRANCE									18.5	X 21.486	23.4	89.3	50.5	
IMMOBILIERE DASSAULT SA									0.13	X 0.081	0.22	0.24	0.49	
IMMOBILIERE HOTELIERE					-29.1	-59.7	-135.9	-37.3	70.3	3.8	X 1.727	8.9	-1.7	

GROUP														
IMMOBILIERE MARSEILLAISE	5.0	?	6.6	5.8	6.6	20.1	7.5	8.4	8.4					Introuvable
INDUS ET FINANC D'ENTREPRISE					2.3	0.9	0.4	1.6	2.7	?	3.7	7.2	6.8	Introuvable
INNELEC MULTIMEDIA SA						1.4	2.2	0.8	0.9	X 1.346	0.7	1.4	1.6	Clôture mars
INST PARTN DU BOIS&DU MEUBL						-5.2	4.0	1.6	0.7	X 0.084	1.7	1.9	X 0.403	
IT LINK							0.5	0.5	-4.1	X -1.976	-0.7	-1.1	-0.5	
LIBERTY SURF GROUP SA									-227.2	X -493.630	-20.3	-50.9	-30.0	
MEDCOST								0.3	0.6	X -2.221	-2.0	-1.5	0.9	
MLPC INTL					1.0	2.2	2.1	1.7	2.3	X -0.43044	0.1	-3.8	-6.9	
MORIA SA							X 2.028	X 7.002	9.1	X 9.407	7.6	4.3		
MUSEE GREVIN	0.2	-1.6	-0.5	-1.5	-0.3	0.0	0.0	-3.3	X 0.190	0.2	0.9	1.4	2.3	
NEXO									1.74	X 1.810	1.65	2.100	2.337	
NOVAMEX									0.81	X 0.352	-0.08	-0.23	X 0.367	
OPTIMS								-3.5	0.3	X -5.011	-1.4	0.0		
ORCHESTRA-KAZIBAO									-2.6	X -4.016	1.8	2.5	3.0	
R2I SANTE SA					0.7	0.0	-0.7		-2.4	-1.9	-5.3	-6.4	-6.1	Date clôture change
RUE IMPERIALE				36.4	56.0	35.8	43.6	115.4	57.0	X 181.189	50.0	23.0		
SCHAEFFER-DUFOUR			2.3	0.7	-11.2	0.0	-9.3	2.3	0.8	X - 2.212	0.8	1.5	2.0	
SICAL						3.4	-0.8	4.6	2.8	1.2	X 9.573	3.1	3.6	
SILICON ON INSULATOR TECHS							-11.3	-7.9	4.1	X 21.962	0.3	-35.8	-10.6	(soitec) Clôture mars
SOAEO-OXYGENE & ACETYL EX-OR	17.2	25.2	26.9	25.9	27.6	27.4	39.4	28.0	33.4	X 36.612	39.6	40.4	43.4	
SOMFY SA	-2.2	X O.710	35.1	35.9	25.8	26.8	11.8	39.0	48.4	57.2	64.4	64.6	74.2	
SOREFICO COIFFURE									-2.2	X -3.910	-3.6	-2.0	-0.6	
STE FRANCAISE PAPIERS PEINT									-1.37	-3.75	-5.51	X -10.902	-9.71	
THERMOCOMPACT SA									1.2	X 0.819	-0.1	-0.8	-3.1	
TITUS INTERACTIVE SA					0.1	1.4	X -2.94314	X 1.36059	-27.6	-47.4	-13.7	-103.5		
TSA TP 83				-413.0	-560.5	-340.0	-170.5	264.3	394.0	X 286.000	373.0	26.0	X - 636.000	(thomson)
ULRIC DE VARENS									1.96	X 2.569	3.93	2.30	2.41	
VINCI	-4.1	13.2	12.8	1.9	16.8	-35.8	-40.1	13.0	2.8	?	143.1	179.4	219.8	(bizarre)

CONSTRUCTION														
WESTERN TELECOM SA									-21.7	X -4.065	0.1	-0.9	X -0.945	

Il apparaît que 46 « trous » sur 73 correspondent à l'année 2001 (63 %). Est-ce le passage à l'euro qui a eu des conséquences sur la tenue et la mise à jour des bases de données ?

Les données concernant les sociétés INDUS ET FINANC D'ENT et IMMOBILIERE MARSEILLAISE n'ont pas été retrouvées, et n'ont pas pu être complétées. Les données de VINCI Construction qui est une société du groupe VINCI (déjà présente dans la base) n'ont pu être retrouvées. Il a été décidé de retirer cette entreprise de l'étude.

Par ailleurs, certains « trous » correspondent à des changements de date de clôture des sociétés. C'est par exemple le cas de la société R2I Santé dont l'exercice 1999-2000 dure 15 mois. Dans une telle situation, l'absence de donnée n'est pas anormale.

Ces recherches ont également permis de compléter le résultat de 2004 lorsqu'il était absent dans la base Global (donnée récente). Au total, 73 données sont récupérées, et le nombre d'entreprises étudiées passe de 1010 à 1009.

b. Correspondance des données pour les variables IB et AT

Suite à la première correction des variables IB manquantes, la correspondance entre les variables IB et AT présentes dans la base a été testée. Il est apparu que pour 53 sociétés, les données manquantes correspondent à des années où le résultat de l'entreprise est connu, mais pas son total actif, ou inversement. Ces données ont donc été complétées en retrouvant l'information dans les rapports annuels disponibles sur Internet ou dans la base Thomson Worldscope.

Le détail de ces données manquantes est fourni dans les Tableau 110 et Tableau 111 ci-dessous :

Tableau 110 : Données manquantes identifiées par confrontation (IB manquant et AT présent)

Company Name	IB1992	IB1993	IB1994	IB1995	IB1996	IB1997	IB1998	IB1999	IB2000	IB2001	IB2002	IB2003	IB2004	Commentaire
AB SOFT SA						? 0.80	-1.09	0.06	-1.64	-2.47	1.69			
ACCESS2NET									X -1.868	X -3.491	-1.07	-0.59	-0.77	
ALAIN AFFLELOU SA									X 12.636	X 14.929	18.62	22.63	22.82	
AUFEMININ.COM									X -7.618	-4.00	-0.91	0.65	1.86	
AUTOROUTES PARIS-RHIN RHONE										X 68.797	74.01	102.11	148.43	
BONDUELLE						9.66	11.39	13.87	X 15.714	24.90	45.22	39.92	38.28	
BUSINESS INTERACTIF									X 0.490	-13.19	-5.88	0.19	0.57	
CFCAL-ALSACE LORRAINE BANQUE					(2.2108)	? 2.3179	? 2.4409	? 2.5833	3.75	4.43	4.44	5.42	34.49	
CIE FRANCAISE MOULAGES METAL						(2.93)	? 2.98	? 4.04	1.10	-5.21	-9.89	0.06	0.47	Cf2m
CITEL SA						? -0.06	? 0.73	? 0.32	0.61	-0.06	0.28	-1.20	-0.07	
COMPAGNIE DES ALPES				8.77	8.37	10.88	13.44	16.98	20.51	23.32	X 21.506 (12 mois)	29.07	29.00	Changement date clôture
CRCAM ALPES-PROVENCE								?	45.93	50.45	62.18	69.92	78.68	
CRCAM NORD					?	?	?	?	41.56	100.74	102.41	126.69	140.94	
CYBER PRESS PUBLISHING						(0.89)	(1.26)	? -2.85	-5.14	-5.03	-12.79	4.41	-5.53	
DAMARTEX						?	?	?	48.43 ou -9.088	? ou -12.670	64.39 ou -2.755	2.24	10.33	Étrange
EURAZEO	? 70.9697 ignorer	41.45 (106.86585 ?)	72.56	111.65	141.01	103.34	127.30	344.31	163.63	583.09	42.71	20.01	13.47	
FIDUCIAL REAL ESTATE					0.69	0.55	0.48	3.00	1.46	X 3.617	4.36	4.55	4.49	
FLIP TECHNOLOGY									X -0.247	X -0.396	X -0.022	X 0.135	X -0.037	
FONCIERE MASSENA	2.51	2.95	3.29	3.00	5.53	5.64	1.88	1.49	-0.36	-0.36 (?)	13.29	16.03	X 4.838	
FRANCO BELGES							? 1.78579	? 1.94896	1.37	X 1.011	X -1.039	X 0.313	X 0.545	
GINGER									X 2.845	X 4.579	4.79	-0.48	1.32	
IMMEUBLES DE FRANCE					? 10.8804	? 14.3786	? 261.124	? 24.761	18.53	21.49	23.39	89.28	50.49	imfr
IMMOBILIERE DASSAULT SA								? 0.62555	0.13	0.08	0.22	0.24	0.49	
INDUS ET FINANC D'ENTREPRISE					2.27	0.89	0.39	1.63	2.70	?	3.70	7.20	6.78	
KALISTO ENTERTAINMENT SA						(1.24)	0.29	1.73	-26.68 (-28.11)					
LCX-LEBLANC CHROMEX						0.56	1.14	.	-0.22	X -0.836	-2.19	0.10	0.24	date clôture. chang
MALTERIES FRANCO-BELGES		-2.25		? -4.16242	6.05	5.10	4.45	2.69	1.04	0.01	5.72	12.39	9.21	Changement

														clôture
MILLIMAGES									X 3.168	X 4.968	-0.70	-1.81	-16.52	
NOVAMEX						? 1.07	? 0.86	? 0.91	0.81 (0.65 ?)	0.35	-0.08	-0.23	0.37	
ORAPI									X 0.617	X 0.941	0.97	0.11	-0.42	
ORCHESTRA-KAZIBAO								? -0.31	-2.59	-4.02	1.79	2.50	3.03	
SOGEPAG					3.53	2.48	? 2.48023	? 2.21399	? 2.324					
SOREFICO COIFFURE						(-1.96)	? -1.95	? -2.93	-2.17	-3.91	-3.56	-2.01	-0.59	
SPORT ELEC SA						(0.493)	(-0.0649)	? -0.006	X 0.392	0.49	0.46	0.34	0.08	
SQLI									0.049	X -7.839	-9.34	1.01	1.95	
STE DES AUTOROUTES DU NORD									X 34.657	X 45.895	11.17	41.86	97.27	(sanef)
SWORD GROUP										X 1.203	3.66	4.89	7.90	
THERMOCOMPACT SA						? 0.904	? 1.219	? 1.263	1.25	0.82	-0.12	-0.79	-3.11	
TOFINSO SDR								? 3.134	X 6.659	X 8.475	4.90	X 3.768		
VINCI CONSTRUCTION	-4.06	13.18	12.83	1.94	16.77	-35.83	-40.13	12.99	2.77	?	143.09	179.37	219.81	
WESTERN TELECOM SA						(0.039)	? -1.956	? -5.536	-21.66	-4.07	0.06	-0.94	-0.95	

En cherchant à compléter les données manquantes, des différences notables ont été constatées entre les bases Thomson Worldscope, et Global Vantage. Ainsi, pour la société Damartex, les données fournies dans la base Global entre 2000 et 2002 sont très différentes de celles issues de la base Thomson Worldscope. En l'absence d'explication, la base n'a pas été modifiée. Le résultat et le total actif de Eurazeo fournis par Thomson Worldscope s'écarte de beaucoup des données de Global Vantage. Pour des problèmes de cohérence dans les ordres de grandeur, l'année 1992 a été ignorée plutôt que de compléter la donnée manquante.

Il apparaît qu'à plusieurs reprises, les données manquantes correspondent à un changement de date de clôture des comptes. C'est par exemple le cas de la compagnie des Alpes. Dans le cadre d'une étude sur les seuils, il a semblé souhaitable de compléter la base avec le résultat sur les 12 derniers mois lorsque celui-ci est disponible.

Au total 80 données sont récupérées.

Tableau 111 : Données manquantes identifiées par confrontation (AT manquant et IB présent)

Company Name	AT92	AT93	AT94	AT95	AT96	AT97	AT98	AT99	AT2000	AT2001	AT2002	AT2003	AT2004	Commentaires
ACCESS2NET									X 10.583	7.23	7.03	6.29	5.76	
ALAIN AFFLELOU SA									190.898	206.33	210.62	325.41	299.01	
CFCAL-ALSACE LORRAINE BANQUE					(424.6574)	443.17	470.29	531.17	550.32	593.25	624.71	655.06	703.02	
CIE FRANCAISE MOULAGES METAL						(69.20)	71.70	83.22	86.48	76.98	68.41	57.00	58.25	
CYBER PRESS PUBLISHING						(4.57)	(12.2)	13.31	30.39	23.84	15.58	14.30	10.58	
DAMARTEX						601.93 ?	657.79 ?	655.09 ?	748.70 ?	781.87 ou 294.466	781.94 ou 238.909	239.95	275.32	Étrange
EPC GROUPE			120.99	133.27	146.39	155.87	150.85	173.02	195.48	213.30	214.65	233.63	X 244.504	
ETUDES REALISATNS MOULES OUT							15.34	19.58	23.96	X 22.106	20.07	20.57	X 24.175	
EURAZEO	693.41 (1899.941 ?) supprimer	762.53 (2044.868?)	2 079.38	2 392.37	2 668.35	2 577.66	2 595.21	3 237.75	3 794.43	2 864.63	2 651.47	2 994.25	4 199.93	
EUROCRYSTAL							X 0.32219	X 0.59768	0.79	1.49	2.10	2.89	3.25	
FLIP TECHNOLOGY									4.97	4.507	4.013	3.129	2.342	
FONCIERE MASSENA	24.93	27.85	27.91	27.52	30.81	32.42	18.08	18.03	17.83	18.98 date ?	19.99	24.91	68.22	
FRANCO BELGES (participations)							39.54	43.23	41.69	39.144	19.174	11.842	12.913	
GINGER									X 52.879	88.24	166.16	183.66	184.39	
GRAND LIVRE DU MOIS					75.62	72.90	70.08	81.47	43.72	X 34.502	32.68	34.32	X 22.102	
GROUPE JAJ				18.68	Date change	27.71	23.66	20.69	20.54	20.59 ou 29.309 ?	33.09	37.31	X 38.369	
IMMOBILIERE HOTELIERE GROUP					798.61	834.87	437.99	309.57	197.16	228.45	X 222.029	170.49	168.86	
INST PARTN DU BOIS&DU MEUBL						32.53	37.01	50.22	54.45	54.24	56.25	53.95	X 52.541	
KALISTO ENTERTAINMENT SA						(3.8)	7.31	21.03	? 20.01					
MALTERIES FRANCO-BELGES		68.01	Date change	70.65 (84.12133 ?)	75.43	77.62	72.24	89.81	109.04	23.95	227.13	207.88	98.40	
MILLIMAGES									X 21.901	50.86	75.42	80.46	107.61	
MUSEE GREVIN	3.96	13.98	12.62	11.46	10.24	9.05	8.63	8.06	? 10.99	15.62	9.04	7.78	9.83	
NEXO									6.96	8.05	9.38	X 11.809	X 12.517	
NOVAMEX						12.68 (6.25 ?)	14.90 (12.58 ?)	? 14.85	15.67	25.05	26.59	21.44	X 19.917	
OPTIMS								12.12	27.82	X 25.342	24.17	29.97		
ORAPI									X 6.454	7.32	13.26	17.13	21.42	
SOGEPAG					32.28	36.95	? 38.45113	41.84	? 68.424					
SOREFICO						(17.84)	15.72	20.23	18.30	16.04	13.80	13.08	14.77	

COIFFURE													
SPORT ELEC SA						(4.781)	(2.999)	3.77	4.35	4.93	4.28	4.21	4.46
SQLI									X 28.573	22.67	17.16	20.61	23.79
STE DES AUTOROUTES DU NORD									X 5 787.086	5 662.32	5 665.94	5 707.16	5 831.65
TOFINSO SDR								143.04 (?)	X 446.468	463.52	445.20	445.789	
TSA TP 83				13 402.12	13 021.21	12 846.19	5 469.69	6 792.53	7 301.00	9 790.00	9 714.00	9 280.00	X 8 525.000
WESTERN TELECOM SA						(2.929)	15.02	20.23	7.57	4.34	3.02	2.27	X 2.204

Les AT des sociétés Damartex, Nexo et Groupe JAJ fournis par la base Global Vantage sont très différents des données fournies par Thomson Worldscope. En l'absence d'explication, les données n'ont pas été modifiées.

Pour la société Malteries, le résultat de l'année 1995 est complété à partir de la base Thomson Worldscope, mais un écart est constaté sur la valeur du total actif. Il semble que si l'on complète la donnée manquante à partir de la base Thomson Worldscope, alors par souci d'homogénéité, il faut également modifier la valeur du total actif figurant dans la base Global.

Pour la société Novamex, les résultats de 1997 et 1998 sont complétés à partir de la base Thomson Worldscope, mais un écart est constaté sur la valeur du total actif. Par souci d'homogénéité, il faut également modifier la valeur du total actif figurant dans la base Global.

Au total, 53 observations sont récupérées.

5. Valeurs extrêmes

Dans la mesure où l'étude s'intéresse essentiellement à la distribution de la variable résultat réduit par une variable de taille, une attention particulière est portée à certaines valeurs très extrêmes. Sont examinés successivement les valeurs extrêmes des ratios $IB/AT_{(n-1)}$ et IB/(AT-IB).

a. Valeur extrême du ratio $IB/AT_{(n-1)}$

Sont appelées valeurs « très extrêmes » les valeurs supérieures à 1 ou inférieures à -1. En effet, cela signifie que l'entreprise réalise des bénéfices (pertes) supérieur(e)s au montant de l'actif. Ces valeurs extrêmes peuvent avoir deux causes :

3. une erreur dans la base, qu'il est utile de corriger.

4. une situation financière extrême justifiée par une conjoncture économique spécifique. Dans ce dernier cas, aucune correction ne s'impose.

Dans le premier cas, la base est corrigée. Dans le second, ces valeurs extrêmes[188] ont seulement été répertoriées.

La revue des cas extrêmes est menée année par année.

Année 1999, deux cas :

Dans la base, les données relatives à la société SYSTRAN sont :

Tableau 112 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 1999 ; Société Systran

	IB			AT			$IB/AT_{(-1)}$	
	1998	1999	2000	1998	1999	2000	99	2000
SYSTRAN SA	0.078	15.661	0.938	2.312	17.959	24.365	6.774	0.052

En 1999, le résultat de Systran est étonnamment élevé. Il est presque sept fois supérieur au Total actif du début de l'année, ce qui conduit les ratios IB/AT(n-1) et IB/(AT-IB) à présenter des valeurs extrêmes. Cette performance est due à une réévaluation des actifs (une reprise sur provision pour dépréciation du fonds de commerce de 14 986 M€ trouvée dans les rapports annuels). C'est une valeur extrême ponctuelle qui n'est pas une erreur et qui n'a pas été corrigée à ce stade.

La société Autoroutes du Sud de la France (ASF) présente également en 1999 une irrégularité.

Tableau 113 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 1999 ; Société ASF

	IB		AT		$IB/AT_{(-1)}$		
AUTOROUTES SUD FRANCE (SD)	IB00	IB01	AT00	AT01	IB00/AT99	IB01/AT00	IB02/AT01
	-0.352	218.400	180.523	11714.598	-0.002	1.210	0.023

La taille de l'entreprise aurait énormément augmenté cette année-là, et la performance comparée à la taille de l'année précédente devient extrêmement élevée. Le problème, c'est que les données de 1999 et 2000 diffèrent entre les deux bases Thomson

[188] Les valeurs extrêmes ont des conséquences importantes sur les mesures paramétriques des irrégularités. Elles font l'objet d'un second examen dans la partie 3.

Worldscope et Global Vantage. Dans la première, le total actif de la société ne présente pas cette énorme progression. La base a été modifiée en se fondant sur les valeurs (moins surprenantes) de la base Thomson Worldscope.

Tableau 114 : Comparaison des AT lus dans les bases Global et Thomson ; année 1999 ; Société ASF

	IB				AT			
AUTOROUTES SUD FRANCE (SD)	IB1999	IB2000	IB2001	IB2002 et suiv.	AT99	AT2000	AT2001	AT2002 et suiv.
Global	0.29499995708	- 0.35199999809	218.4	(…)	189.4490051	180.522995	11714.5	(…)
Thomson	217.90	261.000	Pas de changement		10645.70	11255.70	Pas de changement	

Année 2000, 11 cas :

Tableau 115 : Valeurs extrêmes du ratio IB/AT$_{(n-1)}$; année 2000

	IB			AT				IB/AT$_{(-1)}$		
Company Name	IB1999	IB2000	IB2001	AT98	AT99	AT2000	AT2001	1999	2000	2001
KALISTO ENTERTAINMENT SA	1.725	-26.679		7.3091	21.032	20.01		0.23	-1.26	
WESTERN TELECOM SA	-5.54	-21.657	-4.065	15.017	20.23	7.57	4.344	-0.36	-1.07	-0.53
MEMSCAP SA	-0.18	-2.621	-0.213		2.368	12.227	107.28		-1.10	-0.01
CYBERDECK	-2.16	-5.995	-23.04		4.744	59.41299	40.151		-1.26	-0.38
INFOVISTA SA	-4.46	-6.956	-6.316		5.415	20.146	85		-1.28	-0.31
INTERACTIF DELTA PRODUCTION	-1.88	-3.4611	-4.121	3.2453	2.1566	6.873011	2.254	-0.57	-1.60	-0.59
ORCHESTRA-KAZIBAO	-0.31	-2.5868	-4.016		1.5652	13.53916	57.63199		-1.65	-0.29
NETVALUE SA	-5.5	-20.658	-20.499		11.807	49.79399	26.687		-1.74	-0.41
INTEGRA SA	-6.59	-349.14	-123.99	11.188	159.95	203.697	121.879	-0.58	-2.18	-0.60
INFOSOURCES	-7.16	-73.801	-57.329	13.332	24.999	66.47499	56.536	-0.53	-2.95	-0.86
CYBERSEARCH	-0.57	-4.695	-3.588		1.0862	12.60559	6.469		-4.32	-0.28

Parmi ces 11 entreprises, trois ratios sont inférieurs à -2 (pertes supérieures à deux fois les actifs). Les autres correspondent à des valeurs extrêmes plus modérées.

La société Kalisto réalise en 2000 une forte perte, mais c'est aussi son dernier exercice avant sa disparition. C'est donc une valeur extrême ponctuelle.

Toutes les autres entreprises étaient en 2000 des entreprises jeunes, en phase de développement. Les fortes pertes de 2000 s'accompagnent d'un accroissement de l'AT, ce qui indique que le dénominateur issu d'un exercice antérieur n'est pas un bon indicateur de taille de référence.

Année 2001 et 2002, 4 cas :

La forte perte de la société Coflexip attire l'attention.

Tableau 116 : Valeurs extrêmes du ratio IB/AT$_{(n-1)}$; années 2001 et 2002 ; Société Coflexip

	IB			AT			IB/AT$_{(-1)}$
Company Name	IB2000	IB2001	IB2002	AT2000	AT2001	AT2002	IB2001/AT2000
COFLEXIP SA	222.697	-105320	45.720	1599.192	2135.010	1905.290	-65.858

Après vérification dans les rapports annuels, il s'avère que c'est une double erreur : de signe et d'ordre de grandeur. Le résultat était positif et il manque la virgule avant les trois derniers chiffres. La valeur -105320 a été remplacée par 105.320 dans la base.

Tableau 117 : Valeurs extrêmes du ratio IB/AT$_{(n-1)}$; années 2001 et 2002 ; Société Grande Paroisse

	IB			AT			IB/AT$_{(-1)}$		
Company Name	IB2000	IB2001	IB2002	AT99	AT2000	AT2001	2000	2001	2002
GRANDE PAROISSE	23.858	-1095.248	-33.194	447.674	492.817	429.997	0.053	-2.222	-0.077

La société Grande Paroisse est une entreprise chimique, filiale du groupe Total. Elle fabrique des engrais notamment sous la marque AZF. Le 21 septembre 2001, une explosion à lieu sur le site de Toulouse. La perte phénoménale de 1 095 M€ en 2001 est liée aux conséquences de l'explosion (provisions) et constitue un cas de valeur extrême ponctuel qui ne doit pas être corrigé.

Tableau 118 : Valeurs extrêmes du ratio IB/AT$_{(n-1)}$; années 2001 et 2002 ; Société Gameloft

	IB				AT			IB/AT$_{(-1)}$		
Company Name	IB2000	IB2001	IB2002	IB2003	AT2000	AT2001	AT2002	2001	2002	2003
GAMELOFT	-2.446	-66.216	-23.241	2.357	48.977	16.729	9.932	-1.351	-1.389	0.237

La société Gameloft, spécialisée dans les jeux vidéos, a connu des années difficiles, mais les pertes, bien qu'importantes, sont ponctuelles. Aucune correction n'a été effectuée.

Dans la base, les données relatives à la société GUYANOR (désormais Euro Ressources) sont :

Tableau 119 : Valeurs extrêmes du ratio IB/AT$_{(n-1)}$; années 2001 et 2002 ; Société Guyanor

EURO RESSOURCES SA	1999	2000	2001	2002	2003	2004
IB	-3.658	-16.478	-7.795	-2.138 ou 0.824	-2.229 ou -0.001000	-1.801 ou 12.103
AT	17.485	4.104	0.173	0.148	0.123	10.405
IB/AT(n-1)		-0.942	-1.899	-12.287	-15.003	-14.56

Dès l'année 1999, les pertes de cette société d'exploration minière sont importantes. À partir de 2001, le ratio Résultat (n) sur Actif (n-1) apparaît comme une valeur extrême. Ce qui peut sembler étonnant, c'est que plusieurs années de suite, la société va publier des résultats très fortement négatifs, alors que ses actifs sont proches de zéro. La société n'a pas disparu profitant de nouveaux investissements réguliers. Cette situation peut s'expliquer par l'activité originale de l'entreprise qui n'attend un retour sur ses investissements qu'à long terme. En conclusion, ce cas est atypique et les résultats de la société sont supprimés de la base, de 2001 à 2004[189]. Il faut remarquer au passage que la valeur du total actif de 2002 dans Global ne correspond pas à la valeur trouvée dans le rapport financier. Par ailleurs, le résultat de 2002 dans la base (-2.138) ne correspond pas aux données du rapport annuel (+824). Si la correction avait été effectuée, c'eût été un des deux seuls cas de figure où le résultat eut été supérieur à l'actif total.

Années 2003 et 2004, 5 cas :

Tableau 120 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 2003

	IB			AT				$IB/AT_{(-1)}$		
Company Name	IB2002	IB2003	IB2004	AT2001	AT2002	AT2003	AT2004	2002	2003	2004
CGBI SA	-29.2	-11.544	2.835	47.287	8.585	8.412	4.401	-0.61657	-1.34467	0.337018
BOURSE DIRECT	-1.95	-4.489	2.621	5.979	3.831	3.804	171.168	-0.32597	-1.17175	0.68901
AUREA	-1.9	-0.048	0.404	8.841	0.037	0.056	27.659	-0.21502	-1.29729	7.214285
PERFECT TECHNOLOGIES SA	-3.94	-19.926	-4.74	20.628	18.564	31.22499	12.742	-0.19119	-1.07336	-0.15180

Après vérification, aucune erreur n'a été constatée dans la base concernant ces sociétés. Les pertes 2003 de CGBI et Bourse Direct sont ponctuellement très importantes. C'est également de cas de Highwave Optical Tech en 2004. Elles n'ont pas été corrigées.

Tableau 121 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 2004

	IB		AT		$IB/AT_{(-1)}$	
	IB2003	IB2004	AT2003	AT2004	2003	2004
HIGHWAVE OPTICAL TECH	-14.25	-39.38699	17.87098	14.56099	-0.421236	-2.203962

Les ratios des sociétés Aurea (ou Abertis) et Perfect Technologies SA sont quant à eux soumis à l'influence des décalages temporaires (forts investissements en 2003 et 2004 pour la première, en 2003 pour la seconde). Notons que le résultat (positif) d'Aurea en

[189] Cette situation extrême s'étalant sur plusieurs années, elle ne peut être corrigée par la seule modification de dénominateur.

2004 rapporté à l'actif de l'année précédente conduit à mesurer une valeur positive très extrême, ce qui est rare. L'utilisation du dénominateur IB/(AT-IB) apporte une réponse à ces problèmes.

b. Valeurs extrêmes du ratio IB/(AT-IB)

Réduire la variable résultat par (AT-IB) permet de limiter (ou de « tasser ») les valeurs extrêmes négatives, sans modifier de manière significative la partie centrale de la distribution, sur laquelle figure le seuil d'étude. Le seul cas de figure atypique serait constitué par d'éventuelles valeurs extrêmes positives, mais on sait par ailleurs qu'elles sont rares, et jamais aussi extrêmes que les valeurs négatives.

Ainsi, sur l'échantillon, il n'apparaît que 6 valeurs extrêmes positives (supérieures à 1). À l'exception de Systran en 1999 (déjà évoqué plus haut), les valeurs sont cependant raisonnablement extrêmes puisque jamais supérieures à 2. Dans tous les cas, elles apparaissent comme des événements ponctuels que rien ne permet d'interpréter comme étant des valeurs aberrantes.

Tableau 122 : Valeurs extrêmes du ratio IB/(AT-IB)

	IB				AT				$IB/AT_{(-1)}$				
Company Name	IB2001	IB2002	IB2003	IB2004	AT2001	AT2002	AT2003	AT2004	1999	2001	2002	2003	2004
SYSTRAN									6.817				
FONCIERE MASSENA	-0.36	13.287	16.035	4.838	18.98	19.985	24.913	68.218		-0.018	1.983	1.805	0.076
SOFCO	0.162	3.589	-0.186	-0.405	7.596	5.459	4.29	4.047		0.021	1.919	-0.041	-0.090
PHONE SYSTEMS & NETWORK	-2.33	1.106			3.246	1.7439				-0.417	1.733		
CGBI SA	-8.68	-29.156	-11.54	2.835	47.287	8.585	8.412	4.401		-0.155	-0.772	-0.578	1.809

Ces cas extrêmes font l'objet d'un examen plus approfondi dans la partie 3. Dans le cadre de la partie 2, ils n'ont pas d'impact sur les mesures non paramétriques utilisées.

Annexe 2 : Équations de régressions[190]

Synthèse des (9x13)=117 régressions permettant d'effectuer les (13x12)= 156 mesures par interpolations linéaires

	1992		1993	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 2.0622x + 2.4324	0,4824	y = 1.4338x + 6.2568	0,1887
petits gains uniquement	y = 1.9392x + 1.3311	0,8568	y = 1.323x + 3.3527	0,4891
les deux classes présumées irrégulières	y = 1.9573x + 0.8598	0,8828	y = 1.2927x + 4.1402	0,528
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 2.42e0.2541x	0,6105	y = 5.7464e0.1254x	0,3631
petits gains uniquement	y = 2.3949e0.2443x	0,6661	y = 4.7418e0.1219x	0,5127
les deux classes présumées irrégulières	y = 2.1969e0.2477x	0,7028	y = 5.1791e0.1185x	0,5865
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 9.1293Ln(x) + 0.2	0,5548	y = 6.5832Ln(x) + 4.3511	0,2335
petits gains uniquement	y = 8.0882Ln(x) - 0.075	0,8607	y = 4.8716Ln(x) + 3.3495	0,383
les deux classes présumées irrégulières	y = 8.0554Ln(x) - 0.1593	0,8623	y = 5.0635Ln(x) + 3.8427	0,4672

	1994		1995	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 2.0608x + 4.6622	0,3093	y = 3.0689x + 1.2838	0,5686
petits gains uniquement	y = 1.9257x + 2.1824	0,693	y = 2.9446x - 0.8095	0,9029
les deux classes présumées irrégulières	y = 1.9146x + 2.4695	0,6958	y = 2.939x - 0.6646	0,9035
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 4.6958e0.1761x	0,5254	y = 3.8792e0.2274x	0,7239
petits gains uniquement	y = 4.2488e0.1692x	0,7209	y = 3.6399e0.2203x	0,8805
les deux classes présumées irrégulières	y = 4.2422e0.1693x	0,7206	y = 3.5422e0.2213x	0,8877
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 9.0486Ln(x) + 2.5432	0,35	y = 12.608Ln(x) - 0.5716	0,5631
petits gains uniquement	y = 7.3893Ln(x) + 1.7366	0,5892	y = 11.174Ln(x) - 1.3062	0,7508
les deux classes présumées irrégulières	y = 7.5017Ln(x) + 2.0257	0,616	y = 11.288Ln(x) - 1.0133	0,7686

[190] Détail des régressions dont les R² figurent dans le Tableau 21 situé page 145.

	1996		1997	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 4.0689x + 1.2838	0,7116	y = 3.9865x + 0.2973	0,6398
petits gains uniquement	y = 3.9811x - 1.0081	0,8722	y = 3.8284x - 1.2878	0,8821
les deux classes présumées irrégulières	y = 3.9573x - 0.3902	0,8808	y = 3.8476x - 1.7866	0,89
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 4.1393e0.2593x	0,763	y = 3.0762e0.2905x	0,7653
petits gains uniquement	y = 3.8932e0.2539x	0,8429	y = 3.0539e0.2813x	0,8288
les deux classes présumées irrégulières	y = 3.8444e0.2543x	0,844	y = 2.8116e0.2845x	0,8648
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 17.247Ln(x) - 1.9717	0,7503	y = 16.98Ln(x) - 3.0156	0,6812
petits gains uniquement	y = 15.959Ln(x) - 2.9394	0,8094	y = 15.599Ln(x) - 3.5182	0,8456
les deux classes présumées irrégulières	y = 16.199Ln(x) - 2.3223	0,8512	y = 15.617Ln(x) - 3.4714	0,8456

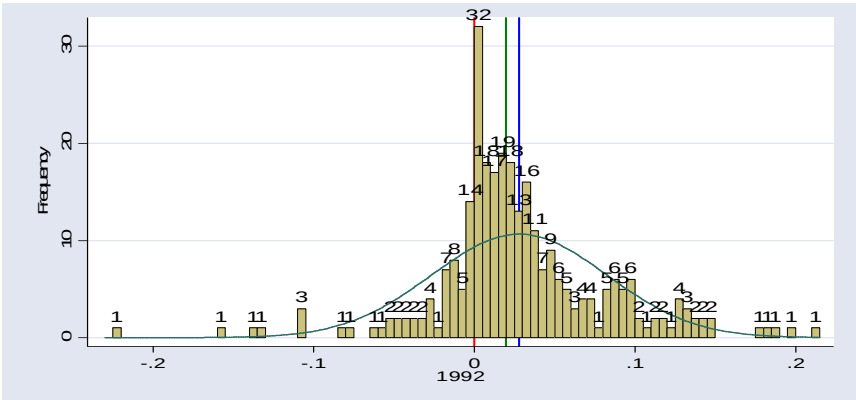
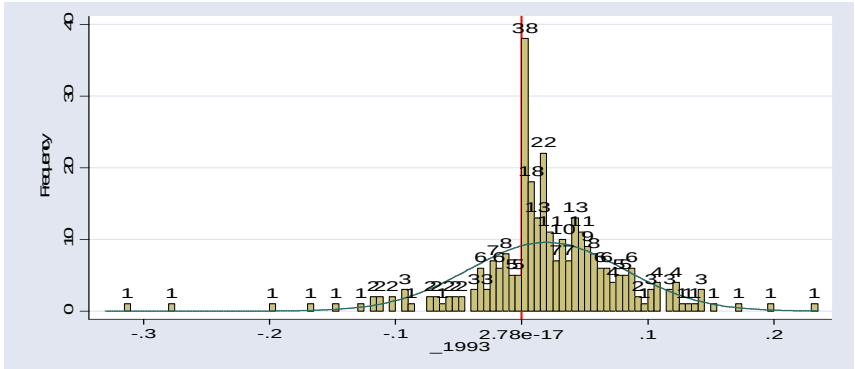
	1998		1999	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 4.8257x - 2.3649	0,634	y = 4.0459x + 1.1892	0,7494
petits gains uniquement	y = 4.7635x - 5.6014	0,7062	y = 4.0541x - 1.4054	0,7376
les deux classes présumées irrégulières	y = 4.7073x - 4.1402	0,7275	y = 3.9878x + 0.3171	0,7896
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 2.4541e0.3229x	0,7093	y = 5.5548e0.2152x	0,7169
petits gains uniquement	y = 2.1807e0.3179x	0,7735	y = 4.8253e0.2153x	0,7135
les deux classes présumées irrégulières	y = 2.2409e0.3168x	0,7743	y = 5.2824e0.2118x	0,7599
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 19.886Ln(x) - 5.3733	0,6318	y = 15.771Ln(x) + 0.0192	0,6682
petits gains uniquement	y = 18.366Ln(x) - 6.8331	0,6062	y = 14.851Ln(x) - 1.3005	0,5716
les deux classes présumées irrégulières	y = 18.791Ln(x) - 5.7396	0,6686	y = 15.303Ln(x) - 0.1372	0,6706

	2000		2001	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 4.9932x - 0.8514	0,6548	y = 4.5554x - 0.4189	0,7071
petits gains uniquement	y = 5.0351x - 3.4135	0,6298	y = 4.5041x - 2.8554	0,7694
les deux classes présumées irrégulières	y = 4.9573x - 1.3902	0,6592	y = 4.4634x - 1.7988	0,7841
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 5.2607e0.2372x	0,7239	y = 4.7621e0.2434x	0,7803
petits gains uniquement	y = 4.6365e0.2375x	0,7178	y = 4.3822e0.2397x	0,8543
les deux classes présumées irrégulières	y = 5.0384e0.2343x	0,7484	y = 4.46e0.239x	0,8554
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 19.5Ln(x) - 2.3504	0,586	y = 18.096Ln(x) - 2.2455	0,6549
petits gains uniquement	y = 18.792Ln(x) - 3.7965	0,5065	y = 16.937Ln(x) - 3.3867	0,6283
les deux classes présumées irrégulières	y = 19.332Ln(x) - 2.4065	0,5782	y = 17.274Ln(x) - 2.5205	0,6773

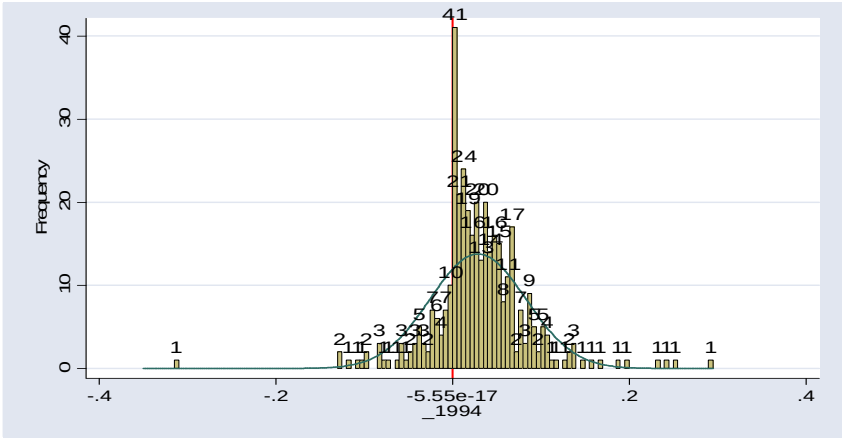
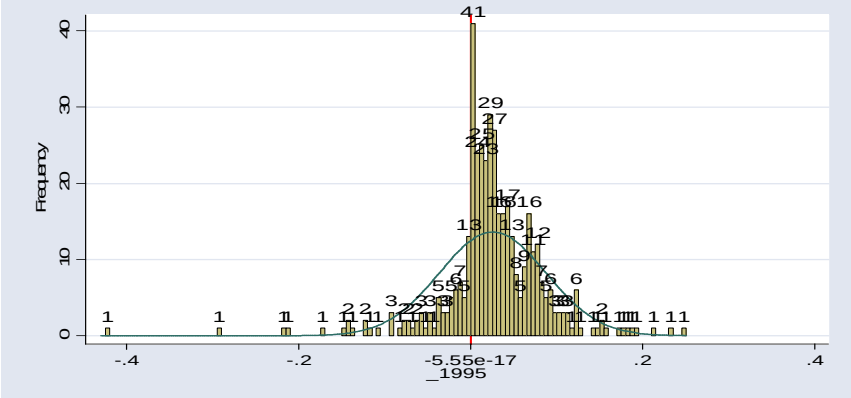
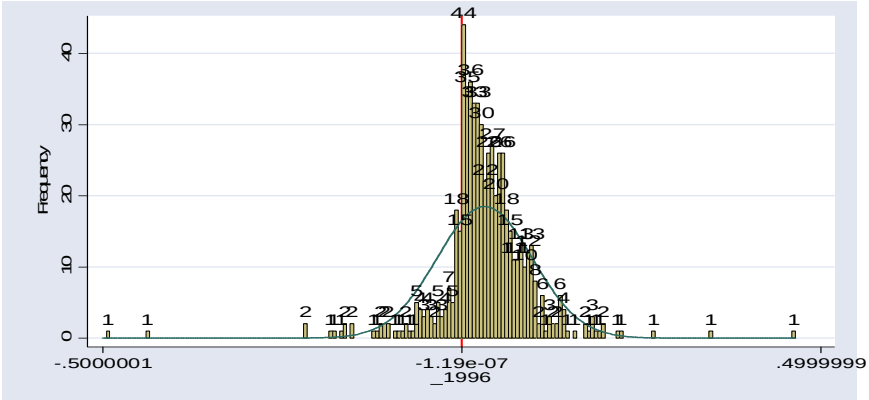
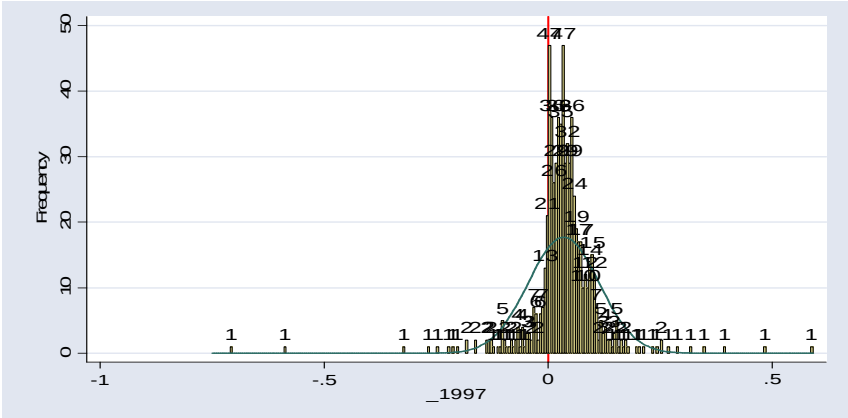
	2002		2003	
interpolation linéaire	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 3.6824x + 4.9865	0,519	y = 3.6973x + 2.4595	0,6714
petits gains uniquement	y = 3.6486x + 2.1351	0,5539	y = 3.6986x + 0.0851	0,6697
les deux classes présumées irrégulières	y = 3.5915x + 3.622	0,5749	y = 3.6402x + 1.6037	0,7066
interpolation exponentielle	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 6.168e0.2109x	0,6791	y = 4.9249e0.2283x	0,6762
petits gains uniquement	y = 5.4957e0.209x	0,7217	y = 4.379e0.2274x	0,6922
les deux classes présumées irrégulières	y = 5.8103e0.2069x	0,7379	y = 4.678e0.2249x	0,7098
interpolation logarithmique	équation	R² expliqué	équation	R² expliqué
petites pertes uniquement	y = 16.241Ln(x) + 1.0927	0,5924	y = 15.209Ln(x) + 0.1949	0,6667
petits gains uniquement	y = 15.018Ln(x) - 0.2143	0,5419	y = 14.348Ln(x) - 1.0006	0,5819
les deux classes présumées irrégulières	y = 15.419Ln(x) + 0.8179	0,6111	y = 14.753Ln(x) + 0.0424	0,6693

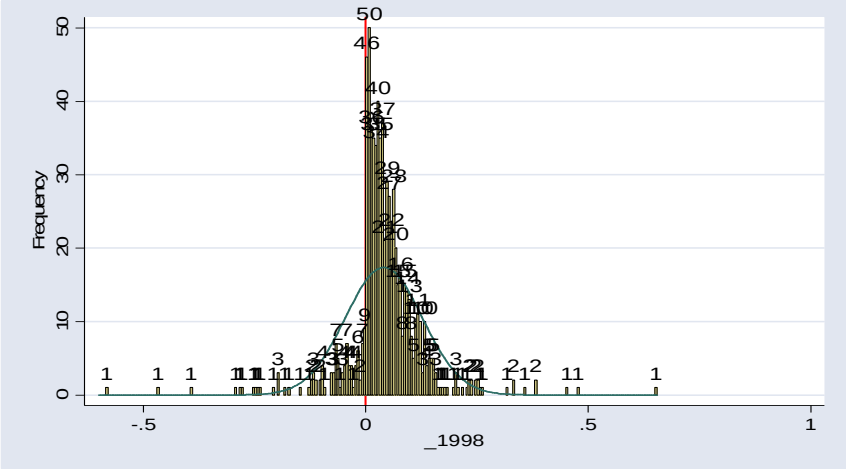
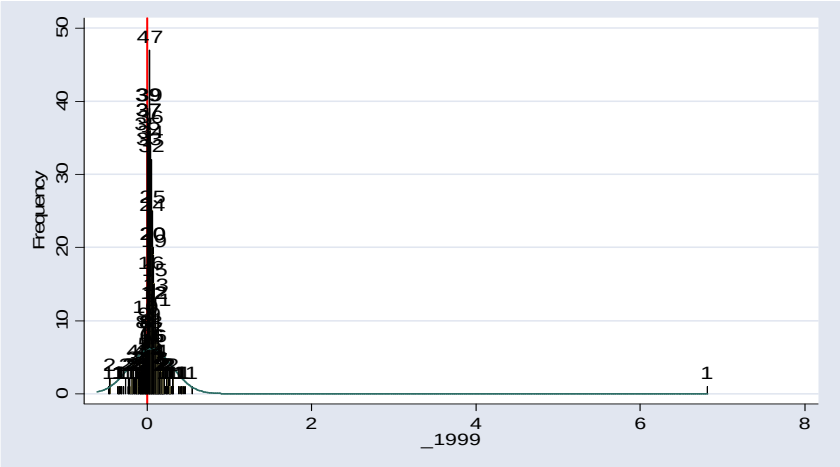
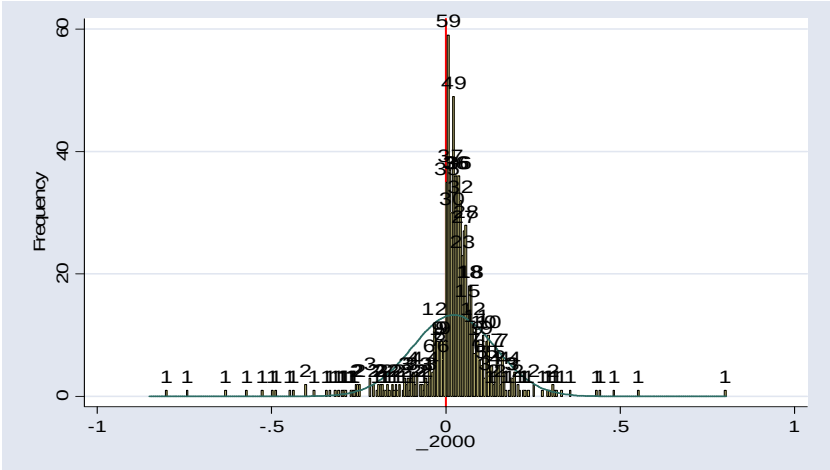
	2004			
interpolation linéaire	équation	R² expliqué		
petites pertes uniquement	y = 4.2176x - 2.9865	0,9447		
petits gains uniquement	y = 4.2743x - 5.3824	0,8521		
les deux classes présumées irrégulières	y = 4.1951x - 3.3232	0,9535		
interpolation exponentielle	équation	R² expliqué		
petites pertes uniquement	y = 2.7611e0.298x	0,8785		
petits gains uniquement	y = 2.1905e0.3026x	0,7585		
les deux classes présumées irrégulières	y = 2.6488e0.2953x	0,902		
interpolation logarithmique	équation	R² expliqué		
petites pertes uniquement	y = 16.647Ln(x) - 4.5163	0,8636		
petits gains uniquement	y = 16.073Ln(x) - 5.8853	0,6957		
Les deux classes présumées irrégulières	y = 16.599Ln(x) - 4.5324	0,8609		

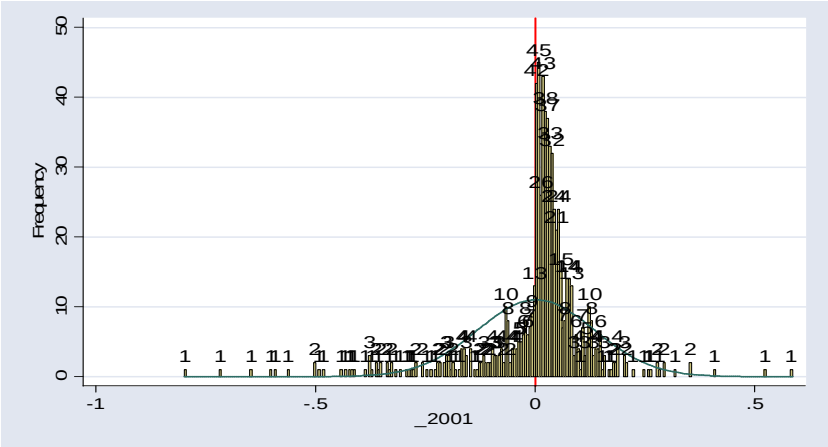
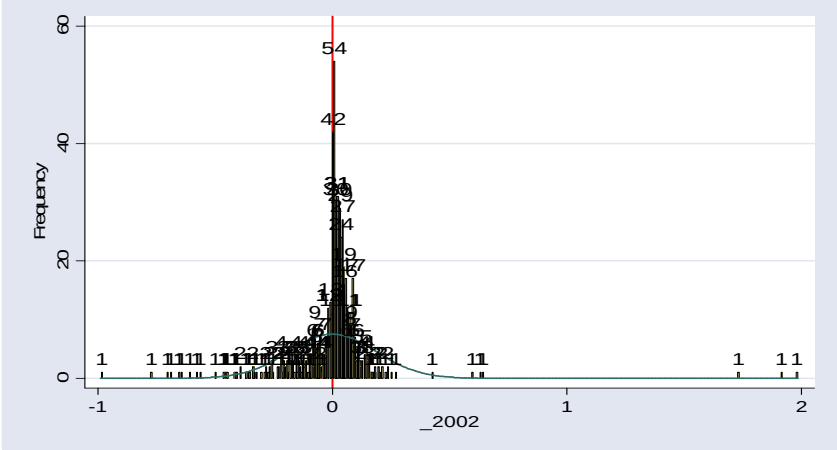
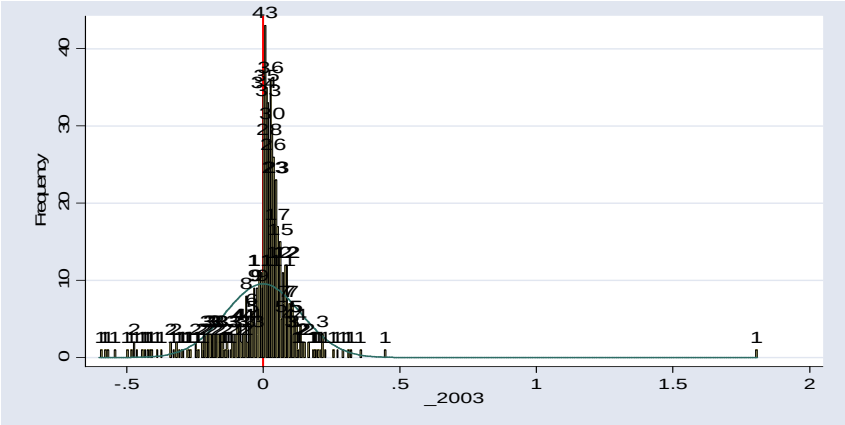
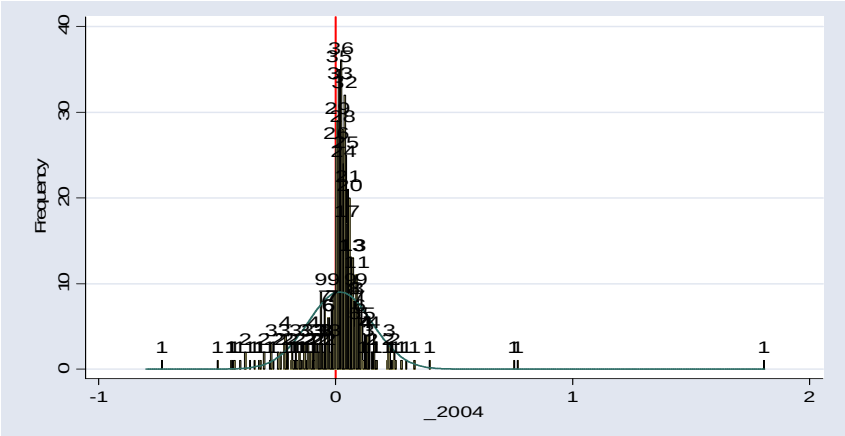
Annexe 3 : Emplacement de l'irrégularité dans la distribution annuelle et choix de la méthode de mesure appropriée[191]

Année	Allure générale de la distribution annuelle	Position du seuil « résultat nul »	Mesure adaptée
	La loi normale est tracée pour faciliter l'analyse graphique et situer approximativement le seuil résultat nul sur la courbe de distribution. Mais elle ne présume pas de la loi réelle de distribution. (Seule hypothèse sous-jacente : la distribution régulière devrait avoir une forme de cloche)	La position est évaluée visuellement ; le seuil est figuré par un trait rouge.	
1992		Pente positive décroissante	Interpolation logarithmique
1993		Pente positive décroissante	Interpolation logarithmique

[191] Détail des analyses graphiques dont les résultats figurent dans le Tableau 33 situé page 163.

1994		Pente positive décroissante	Interpolation logarithmique
1995		Pente positive décroissante	Interpolation logarithmique
1996		Pente positive décroissante	Interpolation logarithmique
1997		Pente positive décroissante	Interpolation logarithmique

1998		Pente positive décroissante	Interpolation logarithmique
1999		Pente positive décroissante	Interpolation logarithmique
2000		Pente positive décroissante	Interpolation logarithmique

2001		Sommet de la courbe	Extrapolation (logarithmique ou linéaire)
2002		Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)
2003		Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)
2004		Pente positive décroissante, proche du sommet	Interpolation ou extrapolation (logarithmique ou linéaire)

Annexe 4 : Observation des intervalles encadrant le seuil, et mesures des irrégularités par moyenne en faisant varier la largeur des intervalles[192]

Largeur	Observations												irrégularité gauche			irrégularité droite		
	Bornes deuxième intervalle gauche			bornes intervalle gauche			bornes intervalle droite			bornes deuxième intervalle droite			MOYENNE			MOYENNE		
	de	à	effectif	de	à	effectif	de	à	effectif	de	à	effectif	(théorique)	écart	écart relatif	(théorique)	écart	écart relatif
0.001	-0.002	-0.001	4	-0.001	0	2	0	0.001	13	0.001	0.002	6	8.5	-6.5	-76.5%	4	9	225.0%
0.002	-0.004	-0.002	10	-0.002	0	6	0	0.002	19	0.002	0.004	19	14.5	-8.5	-58.6%	12.5	6.5	52.0%
0.003	-0.006	-0.003	9	-0.003	0	13	0	0.003	27	0.003	0.006	31	18	-5	-27.8%	22	5	22.7%
0.004	-0.008	-0.004	12	-0.004	0	16	0	0.004	38	0.004	0.008	35	25	-9	-36.0%	25.5	12.5	49.0%
0.005	-0.01	-0.005	13	-0.005	0	21	0	0.005	47	0.005	0.01	36	30	-9	-30.0%	28.5	18.5	64.9%
0.006	-0.012	-0.006	15	-0.006	0	22	0	0.006	58	0.006	0.012	33	36.5	-14.5	-39.7%	27.5	30.5	110.9%
0.007	-0.014	-0.007	15	-0.007	0	25	0	0.007	65	0.007	0.014	40	40	-15	-37.5%	32.5	32.5	100.0%
0.008	-0.016	-0.008	15	-0.008	0	28	0	0.008	73	0.008	0.016	43	44	-16	-36.4%	35.5	37.5	105.6%
0.009	-0.018	-0.009	14	-0.009	0	32	0	0.009	76	0.009	0.018	49	45	-13	-28.9%	40.5	35.5	87.7%
0.01	-0.02	-0.01	13	-0.01	0	34	0	0.01	83	0.01	0.02	55	48	-14	-29.2%	44.5	38.5	86.5%
0.011	-0.022	-0.011	14	-0.011	0	34	0	0.011	89	0.011	0.022	67	51.5	-17.5	-34.0%	50.5	38.5	76.2%
0.012	-0.024	-0.012	12	-0.012	0	37	0	0.012	91	0.012	0.024	77	51.5	-14.5	-28.2%	57	34	59.6%
0.013	-0.026	-0.013	14	-0.013	0	38	0	0.013	101	0.013	0.026	79	57.5	-19.5	-33.9%	58.5	42.5	72.6%
0.014	-0.028	-0.014	14	-0.014	0	40	0	0.014	105	0.014	0.028	92	59.5	-19.5	-32.8%	66	39	59.1%
0.015	-0.03	-0.015	14	-0.015	0	41	0	0.015	109	0.015	0.03	100	61.5	-20.5	-33.3%	70.5	38.5	54.6%
0.016	-0.032	-0.016	15	-0.016	0	43	0	0.016	116	0.016	0.032	112	65.5	-22.5	-34.4%	77.5	38.5	49.7%
0.017	-0.034	-0.017	16	-0.017	0	44	0	0.017	120	0.017	0.034	123	68	-24	-35.3%	83.5	36.5	43.7%
0.018	-0.036	-0.018	16	-0.018	0	46	0	0.018	125	0.018	0.036	136	70.5	-24.5	-34.8%	91	34	37.4%
0.019	-0.038	-0.019	15	-0.019	0	47	0	0.019	134	0.019	0.038	136	74.5	-27.5	-36.9%	91.5	42.5	46.4%
0.02	-0.04	-0.02	17	-0.02	0	47	0	0.02	138	0.02	0.04	147	77.5	-30.5	-39.4%	97	41	42.3%
0.021	-0.042	-0.021	18	-0.021	0	47	0	0.021	145	0.021	0.042	154	81.5	-34.5	-42.3%	100.5	44.5	44.3%
0.022	-0.044	-0.022	18	-0.022	0	48	0	0.022	156	0.022	0.044	157	87	-39	-44.8%	102.5	53.5	52.2%
0.023	-0.046	-0.023	19	-0.023	0	49	0	0.023	163	0.023	0.046	164	91	-42	-46.2%	106.5	56.5	53.1%
0.024	-0.048	-0.024	19	-0.024	0	49	0	0.024	168	0.024	0.048	165	93.5	-44.5	-47.6%	107	61	57.0%
0.025	-0.05	-0.025	21	-0.025	0	49	0	0.025	174	0.025	0.05	172	97.5	-48.5	-49.7%	110.5	63.5	57.5%
0.026	-0.052	-0.026	18	-0.026	0	52	0	0.026	180	0.026	0.052	180	99	-47	-47.5%	116	64	55.2%
0.027	-0.054	-0.027	18	-0.027	0	53	0	0.027	189	0.027	0.054	184	103.5	-50.5	-48.8%	118.5	70.5	59.5%
0.028	-0.056	-0.028	20	-0.028	0	54	0	0.028	197	0.028	0.056	190	108.5	-54.5	-50.2%	122	75	61.5%
0.029	-0.058	-0.029	19	-0.029	0	55	0	0.029	202	0.029	0.058	193	110.5	-55.5	-50.2%	124	78	62.9%
0.03	-0.06	-0.03	20	-0.03	0	55	0	0.03	209	0.03	0.06	197	114.5	-59.5	-52.0%	126	83	65.9%
0.031	-0.062	-0.031	19	-0.031	0	57	0	0.031	220	0.031	0.062	197	119.5	-62.5	-52.3%	127	93	73.2%
0.032	-0.064	-0.032	18	-0.032	0	58	0	0.032	228	0.032	0.064	195	123	-65	-52.8%	126.5	101.5	80.2%
0.033	-0.066	-0.033	18	-0.033	0	58	0	0.033	232	0.033	0.066	199	125	-67	-53.6%	128.5	103.5	80.5%
0.034	-0.068	-0.034	17	-0.034	0	60	0	0.034	243	0.034	0.068	194	130	-70	-53.8%	127	116	91.3%
0.035	-0.07	-0.035	18	-0.035	0	62	0	0.035	256	0.035	0.07	186	137	-75	-54.7%	124	132	106.5%
0.036	-0.072	-0.036	19	-0.036	0	62	0	0.036	261	0.036	0.072	187	140	-78	-55.7%	124.5	136.5	109.6%
0.037	-0.074	-0.037	20	-0.037	0	62	0	0.037	268	0.037	0.074	184	144	-82	-56.9%	123	145	117.9%
0.038	-0.076	-0.038	20	-0.038	0	62	0	0.038	270	0.038	0.076	191	145	-83	-57.2%	126.5	143.5	113.4%
0.039	-0.078	-0.039	21	-0.039	0	62	0	0.039	281	0.039	0.078	185	151	-89	-58.9%	123.5	157.5	127.5%
0.04	-0.08	-0.04	19	-0.04	0	64	0	0.04	285	0.04	0.08	184	152	-88	-57.9%	124	161	129.8%

[192] Détail des mesures effectuées dans le cadre de l'étude de l'influence de la largeur des intervalles de la section VI-B, page 186. Ces mesures sont représentées dans la Figure 65.

0.041	-0.082	-0.041		-0.041	0		0	0.041		0.041	0.082		0					
0.042	-0.084	-0.042	20	-0.042	0	56	0	0.042	299	0.042	0.084	181	159.5	-103.5	-64.9%	118.5	180.5	152.3%
0.043	-0.086	-0.043		-0.043	0		0	0.043		0.043	0.086		0					
0.044	-0.088	-0.044	20	-0.044	0	66	0	0.044	313	0.044	0.088	175	166.5	-100.5	-60.4%	120.5	192.5	159.8%
0.045	-0.09	-0.045		-0.045	0		0	0.045		0.045	0.09		0					
0.046	-0.092	-0.046		-0.046	0		0	0.046		0.046	0.092		0					
0.047	-0.094	-0.047		-0.047	0		0	0.047		0.047	0.094		0					
0.048	-0.096	-0.048	19	-0.048	0	68	0	0.048	333	0.048	0.096	178	176	-108	-61.4%	123	210	170.7%
0.049	-0.098	-0.049		-0.049	0		0	0.049		0.049	0.098		0					
0.05	-0.1	-0.05	19	-0.05	0	70	0	0.05	346	0.05	0.1	174	182.5	-112.5	-61.6%	122	224	183.6%
0.051	-0.102	-0.051		-0.051	0		0	0.051		0.051	0.102		0					
0.052	-0.104	-0.052		-0.052	0		0	0.052		0.052	0.104		0					
0.053	-0.106	-0.053		-0.053	0		0	0.053		0.053	0.106		0					
0.054	-0.108	-0.054		-0.054	0		0	0.054		0.054	0.108		0					
0.055	-0.11	-0.055		-0.055	0		0	0.055		0.055	0.11		0					
0.056	-0.112	-0.056		-0.056	0		0	0.056		0.056	0.112		0					
0.057	-0.114	-0.057		-0.057	0		0	0.057		0.057	0.114		0					
0.058	-0.116	-0.058		-0.058	0		0	0.058		0.058	0.116		0					
0.059	-0.118	-0.059		-0.059	0		0	0.059		0.059	0.118		0					
0.06	-0.12	-0.06	21	-0.06	0	75	0	0.06	406	0.06	0.12	114	213.5	-138.5	-64.9%	94.5	311.5	329.6%
0.061	-0.122	-0.061		-0.061	0		0	0.061		0.061	0.122		0					
0.062	-0.124	-0.062		-0.062	0		0	0.062		0.062	0.124		0					
0.063	-0.126	-0.063		-0.063	0		0	0.063		0.063	0.126		0					
0.064	-0.128	-0.064		-0.064	0		0	0.064		0.064	0.128		0					
0.065	-0.13	-0.065		-0.065	0		0	0.065		0.065	0.13		0					
0.066	-0.132	-0.066		-0.066	0		0	0.066		0.066	0.132		0					
0.067	-0.134	-0.067		-0.067	0		0	0.067		0.067	0.134		0					
0.068	-0.136	-0.068		-0.068	0		0	0.068		0.068	0.136		0					
0.069	-0.138	-0.069		-0.069	0		0	0.069		0.069	0.138		0					
0.07	-0.14	-0.07	23	-0.07	0	80	0	0.07	442	0.07	0.14	115	232.5	-152.5	-65.6%	97.5	344.5	353.3%

Annexe 5 : Mesures des irrégularités par symétrie en faisant varier la largeur des intervalles[193]

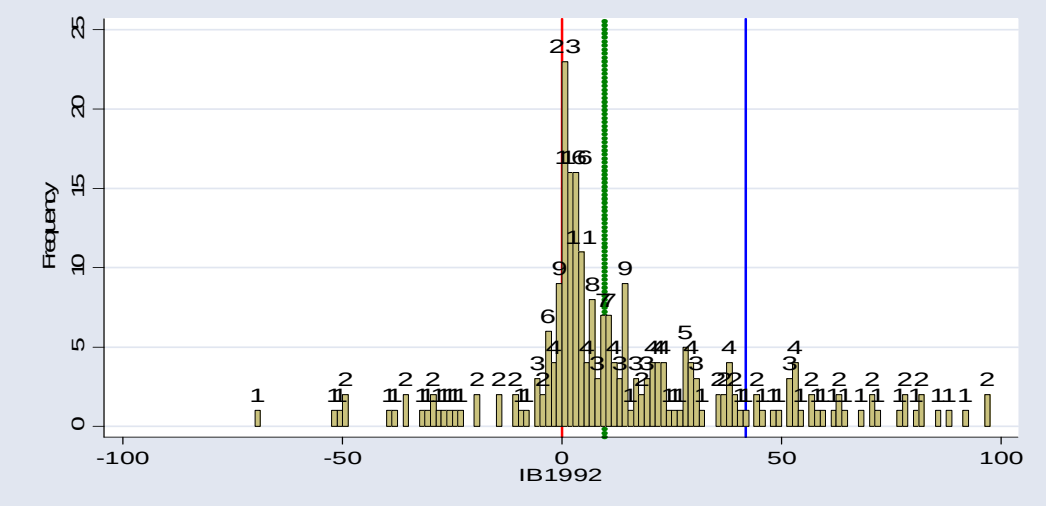
largeur	Observations												irreg inf zéro		irreg sup zéro	
	bornes intervalle inf à zéro			bornes intervalle sup. à 2xmoy			bornes intervalle sup. à zéro			bornes intervalle inf a 2xmoy						
	de	à	effectif	de	à	effectif	de	à	effectif	de	à	effectif	écart	écart relatif	écart	écart relatif
0.001	-0.001	0	2	0.0681	0.0691	4	0	0.001	13	0.0671	0.0681	1	-2	-50.0%	12	1200.0%
0.002	-0.002	0	6	0.0681	0.0701	5	0	0.002	19	0.0661	0.0681	6	1	20.0%	13	216.7%
0.003	-0.003	0	13	0.0681	0.0711	8	0	0.003	27	0.0651	0.0681	12	5	62.5%	15	125.0%
0.004	-0.004	0	16	0.0681	0.0721	11	0	0.004	38	0.0641	0.0681	14	5	45.5%	24	171.4%
0.005	-0.005	0	21	0.0681	0.0731	13	0	0.005	47	0.0631	0.0681	16	8	61.5%	31	193.8%
0.006	-0.006	0	22	0.0681	0.0741	17	0	0.006	58	0.0621	0.0681	20	5	29.4%	38	190.0%
0.007	-0.007	0	25	0.0681	0.0751	22	0	0.007	65	0.0611	0.0681	24	3	13.6%	41	170.8%
0.008	-0.008	0	28	0.0681	0.0761	24	0	0.008	73	0.0601	0.0681	30	4	16.7%	43	143.3%
0.009	-0.009	0	32	0.0681	0.0771	26	0	0.009	76	0.0591	0.0681	37	6	23.1%	39	105.4%
0.01	-0.01	0	34	0.0681	0.0781	29	0	0.01	83	0.0581	0.0681	42	5	17.2%	41	97.6%
0.011	-0.011	0	34	0.0681	0.0791	31	0	0.011	89	0.0571	0.0681	47	3	9.7%	42	89.4%
0.012	-0.012	0	37	0.0681	0.0801	33	0	0.012	91	0.0561	0.0681	49	4	12.1%	42	85.7%
0.013	-0.013	0	38	0.0681	0.0811	35	0	0.013	101	0.0551	0.0681	55	3	8.6%	46	83.6%
0.014	-0.014	0	40	0.0681	0.0821	39	0	0.014	105	0.0541	0.0681	63	1	2.6%	42	66.7%
0.015	-0.015	0	41	0.0681	0.0831	41	0	0.015	109	0.0531	0.0681	70	0	0.0%	39	55.7%
0.016	-0.016	0	43	0.0681	0.0841	43	0	0.016	116	0.0521	0.0681	77	0	0.0%	39	50.6%
0.017	-0.017	0	44	0.0681	0.0851	44	0	0.017	120	0.0511	0.0681	81	0	0.0%	39	48.1%
0.018	-0.018	0	46	0.0681	0.0861	46	0	0.018	125	0.0501	0.0681	89	0	0.0%	36	40.4%
0.019	-0.019	0	47	0.0681	0.0871	51	0	0.019	134	0.0491	0.0681	98	-4	-7.8%	36	36.7%
0.02	-0.02	0	47	0.0681	0.0881	51	0	0.02	138	0.0481	0.0681	102	-4	-7.8%	36	35.3%
0.021	-0.021	0	47	0.0681	0.0891	52	0	0.021	145	0.0471	0.0681	107	-5	-9.6%	38	35.5%
0.022	-0.022	0	48	0.0681	0.0901	54	0	0.022	156	0.0461	0.0681	110	-6	-11.1%	46	41.8%
0.023	-0.023	0	49	0.0681	0.0911	58	0	0.023	163	0.0451	0.0681	120	-9	-15.5%	43	35.8%
0.024	-0.024	0	49	0.0681	0.0921	60	0	0.024	168	0.0441	0.0681	124	-11	-18.3%	44	35.5%
0.025	-0.025	0	49	0.0681	0.0931	63	0	0.025	174	0.0431	0.0681	129	-14	-22.2%	45	34.9%
0.026	-0.026	0	52	0.0681	0.0941	64	0	0.026	180	0.0421	0.0681	138	-12	-18.8%	42	30.4%
0.027	-0.027	0	53	0.0681	0.0951	68	0	0.027	189	0.0411	0.0681	144	-15	-22.1%	45	31.3%
0.028	-0.028	0	54	0.0681	0.0961	74	0	0.028	197	0.0401	0.0681	152	-20	-27.0%	45	29.6%
0.029	-0.029	0	55	0.0681	0.0971	74	0	0.029	202	0.0391	0.0681	154	-19	-25.7%	48	31.2%
0.03	-0.03	0	55	0.0681	0.0981	75	0	0.03	209	0.0381	0.0681	166	-20	-26.7%	43	25.9%
0.031	-0.031	0	57	0.0681	0.0991	79	0	0.031	220	0.0371	0.0681	168	-22	-27.8%	52	31.0%
0.032	-0.032	0	58	0.0681	0.1001	83	0	0.032	228	0.0361	0.0681	176	-25	-30.1%	52	29.5%
0.033	-0.033	0	58	0.0681	0.1011	87	0	0.033	232	0.0351	0.0681	180	-29	-33.3%	52	28.9%
0.034	-0.034	0	60	0.0681	0.1021	89	0	0.034	243	0.0341	0.0681	192	-29	-32.6%	51	26.6%
0.035	-0.035	0	62	0.0681	0.1031	91	0	0.035	256	0.0331	0.0681	204	-29	-31.9%	52	25.5%
0.036	-0.036	0	62	0.0681	0.1041	93	0	0.036	261	0.0321	0.0681	209	-31	-33.3%	52	24.9%
0.037	-0.037	0	62	0.0681	0.1051	95	0	0.037	268	0.0311	0.0681	214	-33	-34.7%	54	25.2%
0.038	-0.038	0	62	0.0681	0.1061	99	0	0.038	270	0.0301	0.0681	227	-37	-37.4%	43	18.9%
0.039	-0.039	0	62	0.0681	0.1071	99	0	0.039	281	0.0291	0.0681	233	-37	-37.4%	48	20.6%
0.04	-0.04	0	64	0.0681	0.1081	100	0	0.04	285	0.0281	0.0681	239	-36	-36.0%	46	19.2%
0.041	-0.041	0	64	0.0681	0.1091	101	0	0.041	292	0.0271	0.0681	247	-37	-36.6%	45	18.2%
0.042	-0.042	0	56	0.0681	0.1101	102	0	0.042	299	0.0261	0.0681	255	-46	-45.1%	44	17.3%
0.043	-0.043	0	65	0.0681	0.1111	102	0	0.043	307	0.0251	0.0681	263	-37	-36.3%	44	16.7%

[193] Détail des mesures effectuées dans le cadre de l'étude de l'influence de la largeur des intervalles de la section VI-B, page 189. Ces mesures sont représentées dans la Figure 68.

0.044	-0.044	0	66	0.0681	0.1121	102	0	0.044	313	0.0241	0.0681	267	-36	-35.3%	46	17.2%
0.045	-0.045	0	67	0.0681	0.1131	104	0	0.045	317	0.0231	0.0681	274	-37	-35.6%	43	15.7%
0.046	-0.046	0		0.0681	0.1141	104	0	0.046		0.0221	0.0681	280				
0.047	-0.047	0		0.0681	0.1151		0	0.047		0.0211	0.0681					
0.048	-0.048	0	68	0.0681	0.1161	108	0	0.048	333	0.0201	0.0681	299	-40	-37.0%	34	11.4%
0.049	-0.049	0		0.0681	0.1171		0	0.049		0.0191	0.0681					
0.05	-0.05	0	70	0.0681	0.1181	109	0	0.05	346	0.0181	0.0681	309	-39	-35.8%	37	12.0%
0.051	-0.051	0		0.0681	0.1191		0	0.051		0.0171	0.0681					
0.052	-0.052	0	70	0.0681	0.1201	109	0	0.052	360	0.0161	0.0681	321	-39	-35.8%	39	12.1%
0.053	-0.053	0		0.0681	0.1211		0	0.053		0.0151	0.0681					
0.054	-0.054	0		0.0681	0.1221		0	0.054		0.0141	0.0681					
0.055	-0.055	0		0.0681	0.1231		0	0.055		0.0131	0.0681					
0.056	-0.056	0		0.0681	0.1241		0	0.056		0.0121	0.0681					
0.057	-0.057	0		0.0681	0.1251		0	0.057		0.0111	0.0681					
0.058	-0.058	0		0.0681	0.1261		0	0.058		0.0101	0.0681					
0.059	-0.059	0		0.0681	0.1271		0	0.059		0.0091	0.0681					
0.06	-0.06	0	75	0.0681	0.1281	114	0	0.06	406	0.0081	0.0681	364	-39	-34.2%	42	11.5%
0.061	-0.061	0		0.0681	0.1291		0	0.061		0.0071	0.0681					
0.062	-0.062	0		0.0681	0.1301		0	0.062		0.0061	0.0681					
0.063	-0.063	0		0.0681	0.1311		0	0.063		0.0051	0.0681					
0.064	-0.064	0	76	0.0681	0.1321	117	0	0.064	423	0.0041	0.0681	396	-41	-35.0%	27	6.8%
0.065	-0.065	0		0.0681	0.1331		0	0.065		0.0031	0.0681					
0.066	-0.066	0		0.0681	0.1341		0	0.066		0.0021	0.0681					
0.067	-0.067	0		0.0681	0.1351		0	0.067		0.0011	0.0681					
0.068	-0.068	0	77	0.0681	0.1361	118	0	0.068	437	0.0001	0.0681	437	-41	-34.7%	0	0.0%
0.069	-0.069	0		0.0681	0.1371		0	0.069		-0.0009	0.0681					
0.07	-0.07	0	80	0.0681	0.1381	118	0	0.07	442	-0.0019	0.0681	442	-38	-32.2%	0	0.0%

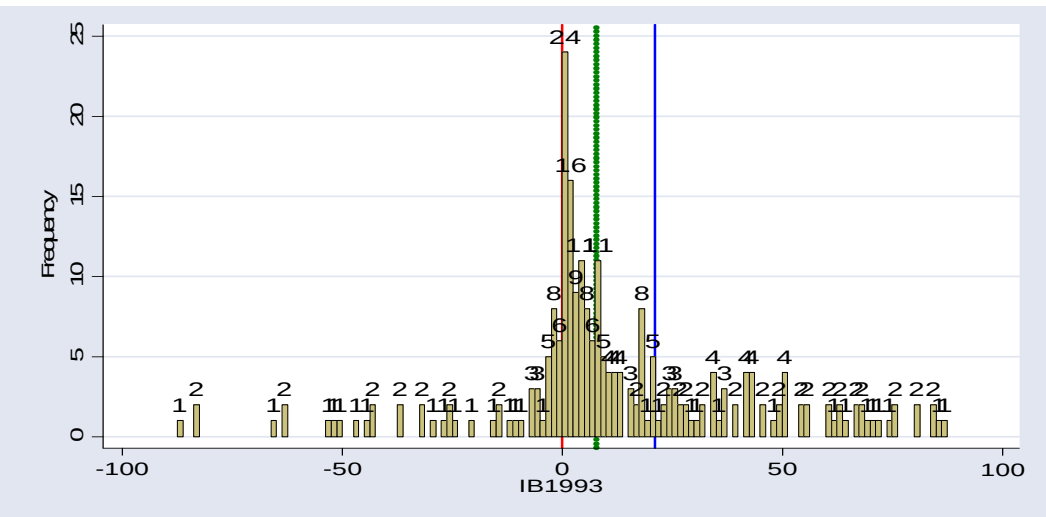
Annexe 6 : Distributions annuelles de la variable résultat non mise à l'échelle[194]

1992

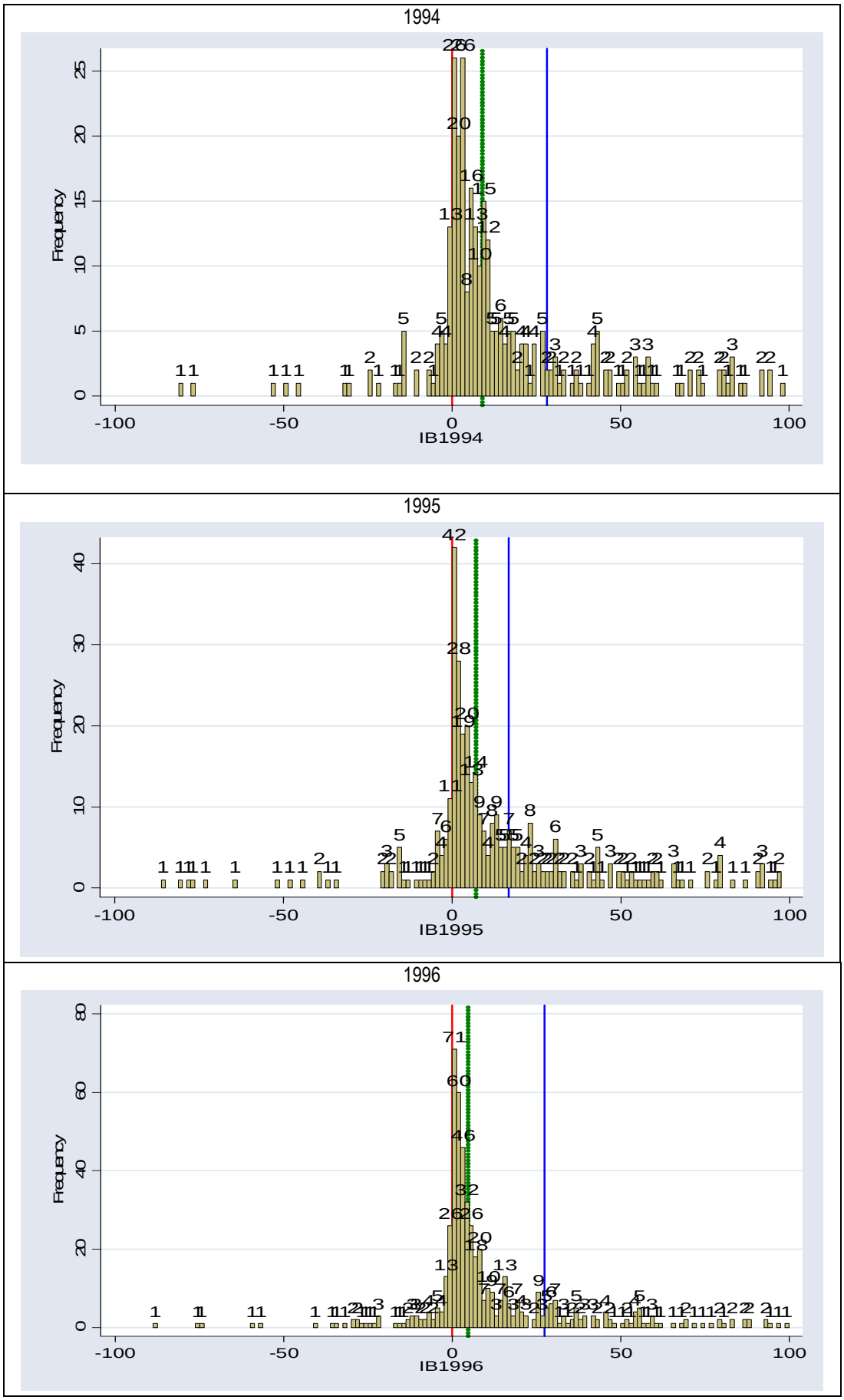


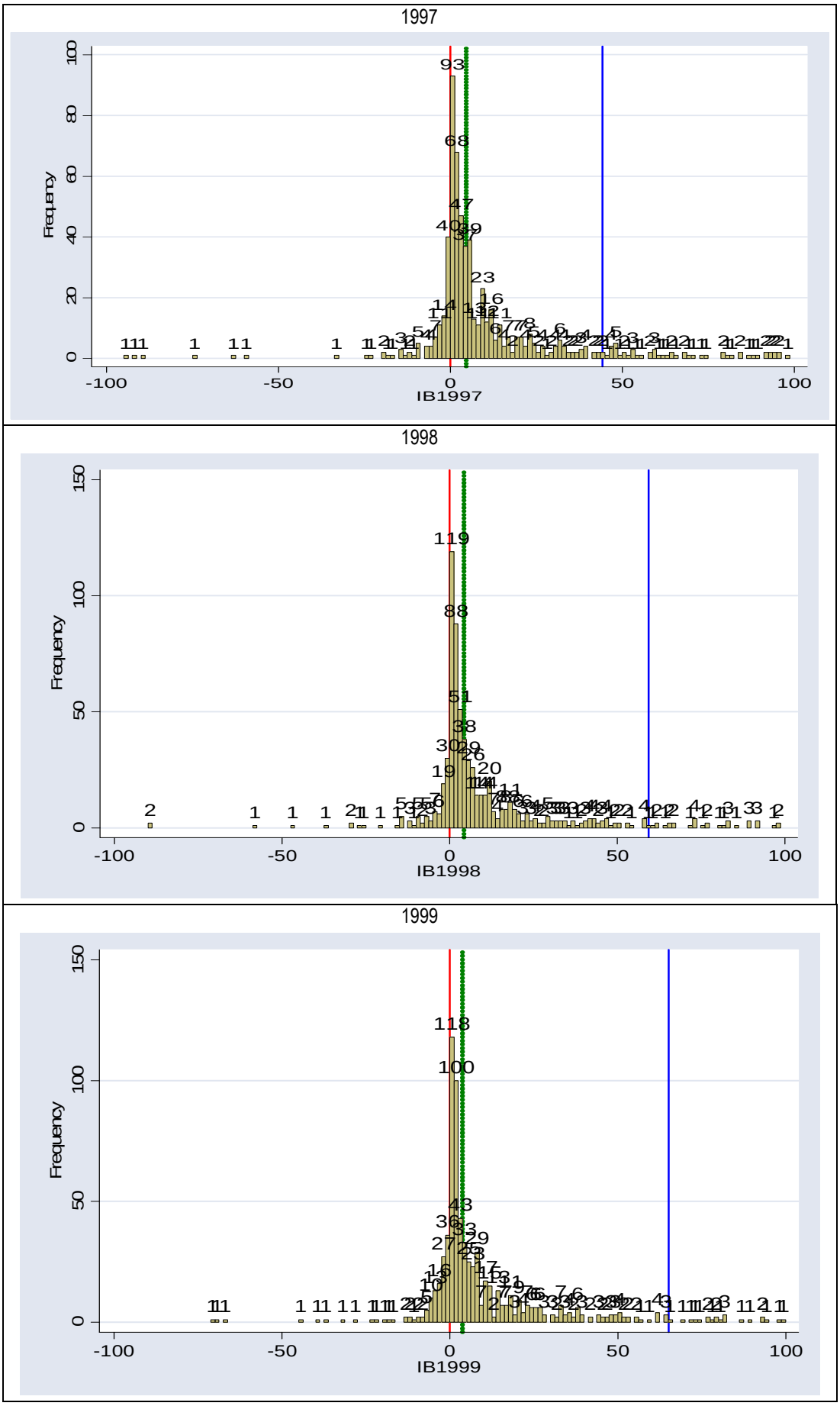
À titre d'exemple, la formule (sous stata) pour l'année 1992 est : histogram IB1992 if IB1992>-100 & IB1992<100, width(1.25) start(-100.0000000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(41.84, lcolor(blue)) xline(9.701,lcolor(green) lwidth(thick) lpattern(dot))

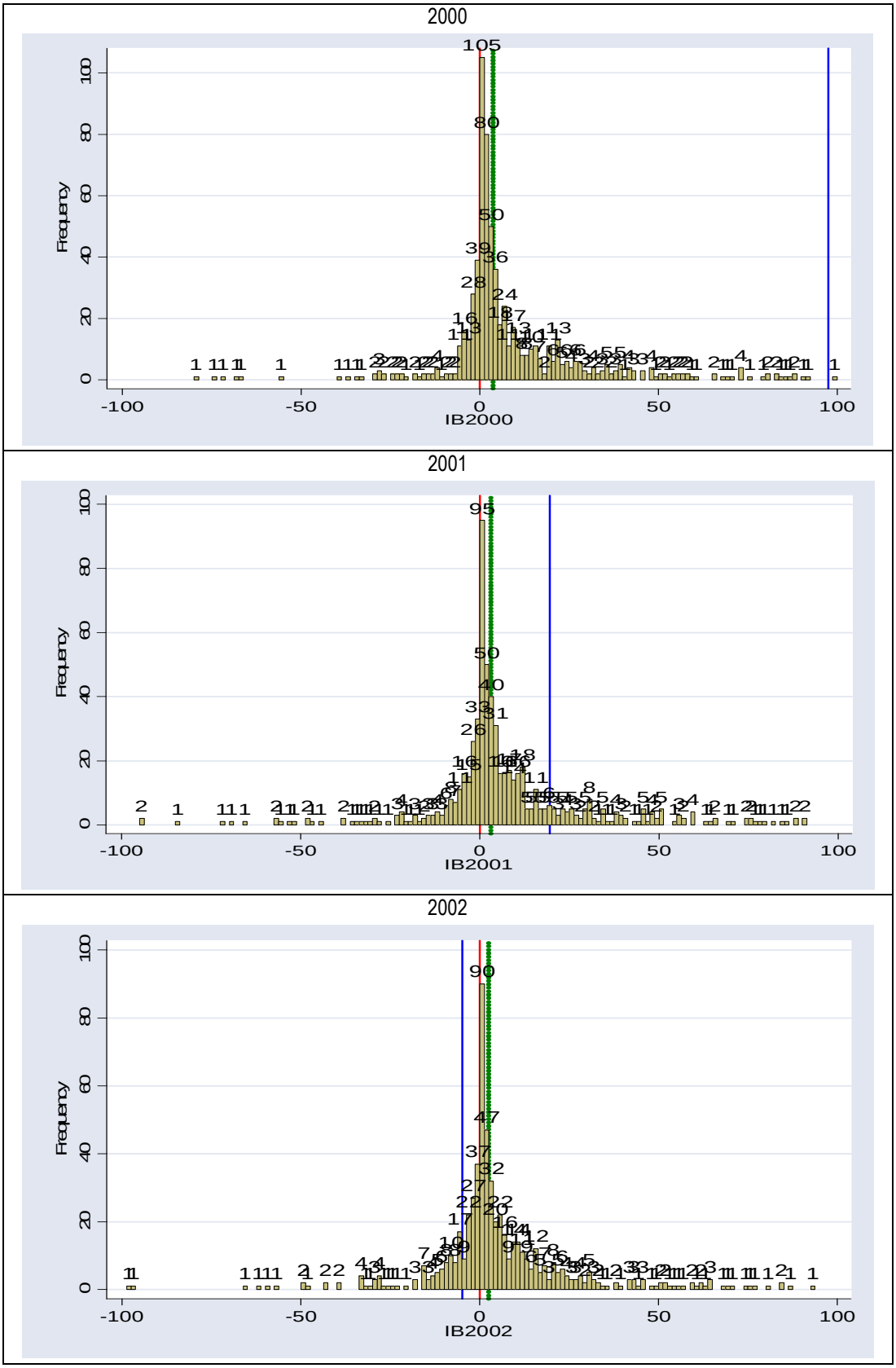
1993

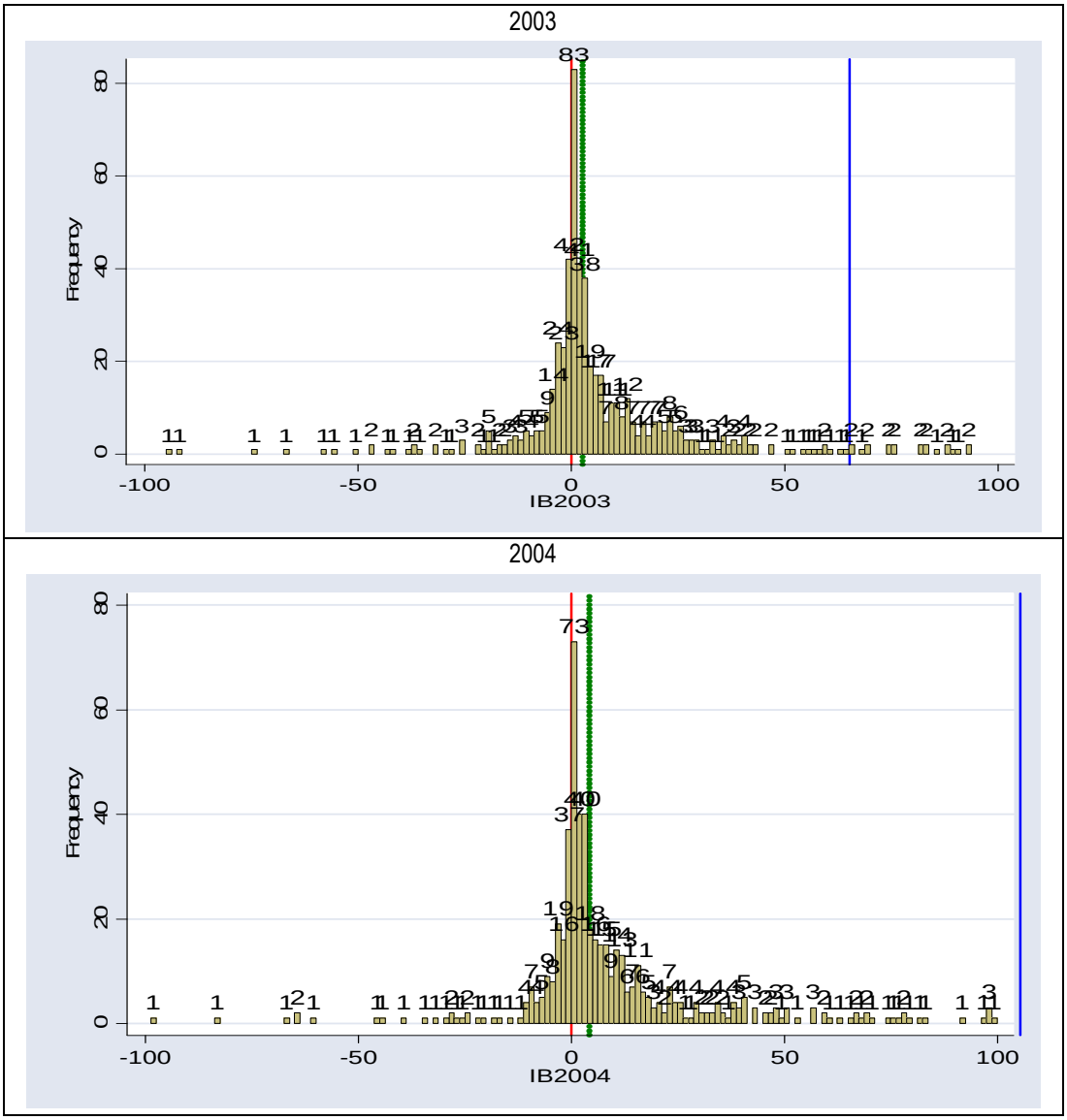


[194] Ces distributions sont commentées dans le cadre de la section VI-C, page 198.


1994
Frequency
IB1994
1995
Frequency
IB1995
1996
Frequency
IB1996


1997
Frequency
IB1997
1998
Frequency
IB1998
1999
Frequency
IB1999


2000
Frequency
IB2000
2001
Frequency
IB2001
2002
Frequency
IB2002


2003
Frequency
0
20
40
60
80
-100
-50
0
50
100
IB2003
2004
Frequency
0
20
40
60
80
-100
-50
0
50
100
IB2004

Annexe 7 : Détail de l'examen de la variable CA dans la base[195]

27 chiffre d'affaires négatifs ou nuls dans la base (en gras dans le Tableau 123) ont été supprimés.

Tableau 123 : Chiffre d'affaires strictement nul ou négatif dans la base

Company Name	CA[Y92]	CA[Y93]	CA[Y94]	CA[Y95]	CA[Y96]	CA[Y97]	CA[Y98]	CA[Y99]	CA[Y00]	CA[Y01]	CA[Y02]	CA[Y03]	CA[Y04]
RUBIS & CIE	**0**	67.87	93.28	121.8	124.8	133.3	184.5	234.9	374.81	376.18	389.64	396.76	413.64
EURO RESSOURCES SA						1.415	0.176	**0**	**0**	0.032	0.4945	0.0903	0.0781
FFP-SOC FONC FINAN ET PARTIC	78.52	**-26.1**	87.47	51.12	27.18	**-62.8**	5.139	7.266	4.907	4.196	**-1.654**	28.687	14.857
FINAXA			22529	23885	30480	0.606	111.1	**-19**	526	33	**-70**	**-46**	**-45**
MINES DE KALI SAINTE THERESE			139.7	108.7	118.3	22.01	2.77	**0**					
SOFIDAV				**0**									
INST PARTN DU BOIS&DU MEUBL						**-3.87**	1.252	0.753	1.5297		1.807	1.427	
ALTAMIR & CIE						0.357	2.464	8.833	5.52	**-5.088**	**-7.987**	6.339	10.934
DARBLAY SA	0.901	0.512	**-0.45**	0.228	0.16	0.942							
EUROTUNNEL GROUPE SA	**0**	**0**	38.32	344.2	604.6	668.4	919.3	968.8	944.11	892.98	896.92	812.75	788.89
LA SAVOISIENNE					**0**								
NICOX SA							2.448	0.227	5.61	**0**	5.468	1.404	1.918
BCI NAVIGATION								**0**	0.7484	0.425	0.692	0.071	
VEOLIA ENVIRONNEMENT								**0**	26394	29127	30079	28603	24673
ORANGE SA									**0**	15087	17085	17941	
SECHILIENNE								**0**	**0**	52.405	43.524	20.264	52.287
AUTOROUTES SUD FRANCE (SD)								**0**	**0**	1929.8	2104.2	2238.7	2389.2

Les variations extrêmes (variation positive supérieure à 10 fois le chiffre d'affaires de l'année précédente ou variation négative inférieure à 90 %) ont été identifiées (en gras dans le Tableau 124).

[195] Détail de l'étude menée dans la section VI-C page 210.

Tableau 124 : Variations extrêmes du chiffre d'affaires

Company Name	CA[Y92]	CA[Y93]	CA[Y94]	CA[Y95]	CA[Y96]	CA[Y97]	CA[Y98]	CA[Y99]	CA[Y00]	CA[Y01]	CA[Y02]	CA[Y03]	CA[Y04]
FRANCE SA	901.9	1061	1125	1189	**10.87**								
COFIFAP			1.092	0.128	0.04	**0.98**							
DEXIA FRANCE SA	3813	4042	3927	4501	949.9	1100	**1 ***						
WORMS & CIE (OLD)	39.07	98.42	120.7	155.8	176.5	**6697**	6173						
COFLEXIP SA	273.1	304.1	309.7	617	743.9	1109	1144	1017	1064.6	**1898820 ***	**1867.8**		
EURO RESSOURCES SA						1.415	0.176			0.032	**0.4945 ***	0.0903	0.0781
AMEC SPIE SA				2630	2357	**116.9**	34.71	**3646**	3752.9	3286.326	3423.5	2454.5	2585.7
WENDEL INVESTISSEMENT	2272	4191	3981	4031	**312.3**	350.1	398.2	442.9	486.52	504.536	552.69	596.73	889.85
CGIP-CIE GEN D'INUST ET PART	2272	4134	3941	3973	**248.9**	281.7	288.5	360.9	431				
TRANSGENE						1.483	9.064	0.647	**2.253 ***	0.66	2.197	2.459	2.514
FINATIS SA					556.4	640.5	**12798**	16383	19955	22881	23728	23868	23856
FINAXA			22529	23885	30480	**0.606**	**111.1**		526	**33 ***			
IMMOBILIERE HOTELIERE GROUP					266.4	352.4	308.3	158	**6.8423**	6.94399		7.494	7.698
PARFINANCE	43.11	47.05	17.44	28.27	32.83	**3.273**							
RALLYE				473.3	523.9	592.4	**12778**	16374	19942	22863	23681	23768	23835
CIE AGRICOLE DE LA CRAU SA					0.802	2.042	12.22	12.49	12.867	12.451	12.638	12.432	**0.195 ***
ALTAREA					2.481	1.44	1.105	1.178	0.9463	0.902	0.978	2.604	**140.39 ***
CARPINIENNE PARTICIPATIONS						0.796	0.499	0.339	0.1732	0.083	**1.021**	2.562	3.57
MAROCAINE (COMPAGNIE)	1.525	1.526	0.911	**0.039**	0.042	0.05	0.041	0.047	0.0485	0.045	0.045	0.043	0.042
FONCIERE EURIS	13.72	**490.3**	463.9	510.4	544	615.1	**12791**	16383	19954	22881	23719	23861	23852
ELECTRICITE & EAUX MADAGASCA			0.304	**3.798**	16.89	26.59	47.48	35.19	35.998	45.308	41.012	40.478	38.322
SOFCO				221.3	244.7	179.2	107.4	**1.097**	3.0786	1.39899	0.219	0.168	0.102
FIDUCIAL REAL ESTATE					1.118	0.966	0.851	**11.73**	14.386		19.861	21.582	25.031
NICOX SA							2.448	**0.227 ***	**5.61**		5.468	1.404	1.918
HIGHWAVE OPTICAL TECH								4.66	**71.912 ***	26.0099	3.297	3.245	3.575

* : cas dont la variation extrême n'est pas confirmée les années suivantes

Certaines variations importantes représentent manifestement des phases de transition. Le nouveau niveau d'activité est confirmé les années suivantes. Ces cas ont donc été conservés. Ont également été conservées les fortes baisses l'année de sortie de cote. Neuf cas ont été examinés de plus près (voir Tableau 125).

Tableau 125 : Sélection des neuf cas dont les variations extrêmes ne sont pas confirmées par la suite

	1997	1998	1999	2000	2001	2002	2003	2004	
COFLEXIP SA		1144	1017	1064.6	**1898820**	1867.8			Il y a manifestement oubli d'une virgule
FINAXA	**0.606**	111.1		526	33				Une recherche sur Internet montre que le CA de 1997 n'est pas représentatif de l'activité de l'entreprise
DEXIA FRANCE SA	1100	**1**							Une recherche sur Internet montre que le CA de 1998 n'est pas représentatif de l'activité de l'entreprise
EURO RESSOURCES SA		0.176			0.032	**0.4945**	0.0903	0.0781	Le résultat à partir de 2001 est non disponible, donc correction inutile
TRANSGENE		9.064	0.647	**2.253**	0.66	2.197	2.459	2.514	Le résultat ne connaît pas de forte variation sur la période. Ni AT, ni VM. La cause de la variation mérite d'être identifiée. IB/CA en 2000 = -13 Ce ratio ne figure donc jamais dans les valeurs proches de l'irrégularité. Il n'apparaît donc pas nécessaire de corriger la base.
CIE AGRICOLE DE LA CRAU SA		12.22	12.49	12.867	12.451	12.638	12.432	**0.195**	L'année 2004 ne voit pas de forte variation du résultat, ni du total actif, ni de la valeur de marché. Seule la variation du CA semble extrême. IB/CA = 3,7 n'est pas pour autant une valeur extrême.
ALTAREA		1.105	1.178	0.9463	0.902	0.978	2.604	**140.39**	L'année 2004 correspond à une forte progression du CA, du résultat, du total actif et de la valeur de marché.
NICOX SA		2.448	**0.227**	5.61		5.468	1.404	1.918	Seul le CA connaît une forte variation en 1999. le ratio est de 15,8. Il n'est donc pas situé sur le seuil d'étude.
HIGHWAVE OPTICAL TECH			4.66	**71.912**	**26.00999**	3.297	3.245	3.575	La variation de la variable CA est corrélée à la variation du total actif et de la valeur de marché.

Le chiffre d'affaires de la société Coflexip en 2001 constitue manifestement une erreur dans la base qui a été corrigée : 1 898,820 au lieu de 1 898 820. Cette correction est importante car un dénominateur extrême aurait conduit le ratio à se situer proche du seuil étudié. Par ailleurs, les chiffres d'affaires de Finaxa et Dexia respectivement en 1997 et 1998 ne sont pas représentatifs de la taille de ces entreprises. Ces deux valeurs qui conduisent à des ratios IB/CA extrêmes sont supprimées. Les autres cas sont, ou corrélés aux variations du total actif et de la valeur de marché (et ne semblent donc pas manifestement irréguliers), ou ne conduisent pas à une valeur extrême après mise à l'échelle. Elles n'ont donc pas été modifiées.

Annexe 8 : Largeurs théoriques en fonction du dénominateur[196]

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	Moyenne des distrib annuelles	Distrib agrégée
IB / CA															
Nombre	288	291	373	418	612	687	736	772	772	673	688	649	597		7556
Iqr	0.06352	0.06138	0.05987	0.06276	0.06582	0.05775	0.65916	0.06631	0.06782	0.07326	0.08211	0.08213	0.07690		0.0679
2IQRn(-1/3)	0.0192	0.0185	0.0166	0.0168	0.0155	0.0131	0.0146	0.0145	0.0148	0.0167	0.0186	0.0190	0.0183	**0,0166**	0.0069
IB / (AT-IB)															
nombre	291	296	376	422	616	702	753	798	791	734	698	659	610		7746
Iqr	0.048	0.052	0.050	0.051	0.054	0.054	0.063	0.063	0.056	0.060	0.064	0.061	0.054		0.05581
2iqrN-1/3	0.0144	0.0155	0.0139	0.0137	0.0128	0.0122	0.0139	0.0136	0.0121	0.0133	0.0144	0.0141	0.0128	**0,0135**	0.00564
IB / VM(-1)															
nombre	0	160	203	291	356	480	494	495	531	650	633	625	582		5500
Iqr	.	0.065	0.041	0.062	0.069	0.076	0.072	0.081	0.096	0.102	0.126	0.154	0.092		0.08753
2iqrN-1/3		0.0239	0.0140	0.0186	0.0194	0.0195	0.0183	0.0206	0.0237	0.0236	0.0294	0.0360	0.0220	**0,0224**	0.00991

[196] Détail des calculs menés dans la section VI-C page 218.

Annexe 9 : Irrégularités mesurées en mettant le résultat à l'échelle par le chiffre d'affaires[197]

Les mesures sont effectuées à partir des distributions annuelles dont les caractéristiques sont récapitulées dans le tableau ci-dessous :

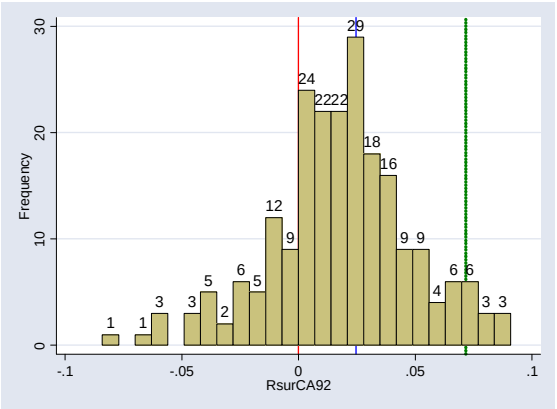
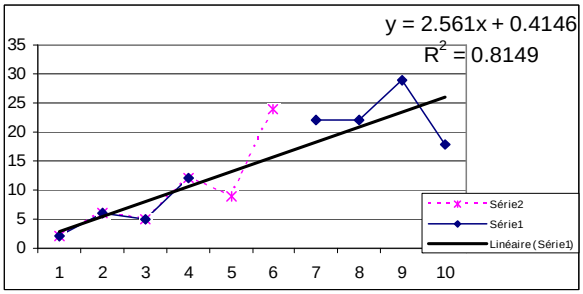
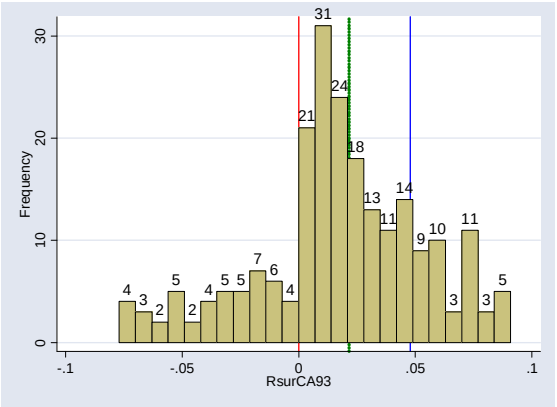
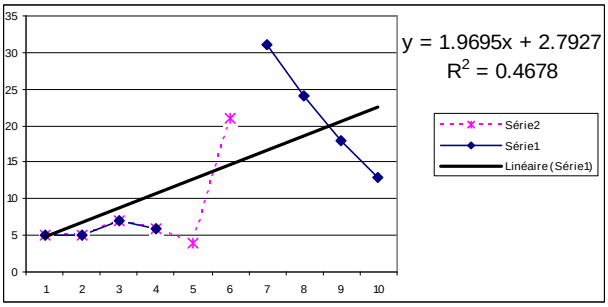
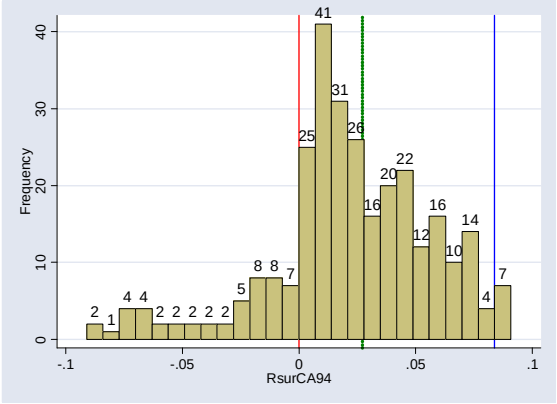
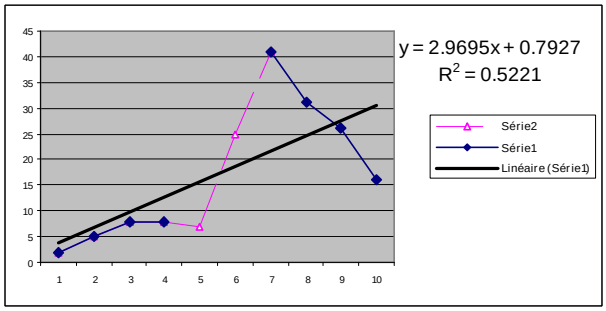
Tableau 126 : Caractéristiques des distributions annuelles de IB mis à l'échelle par le chiffre d'affaires

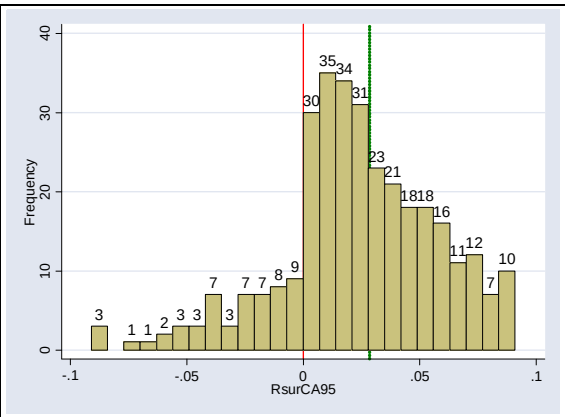
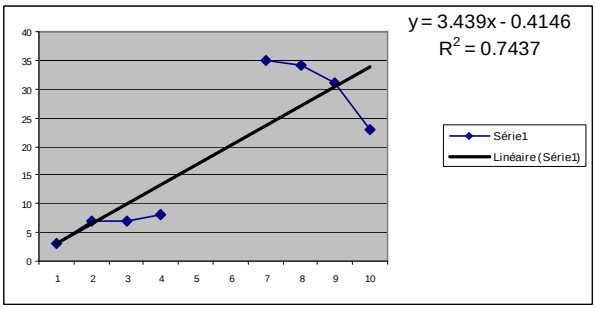
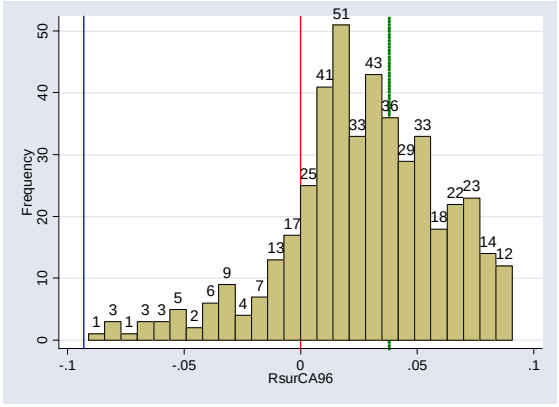
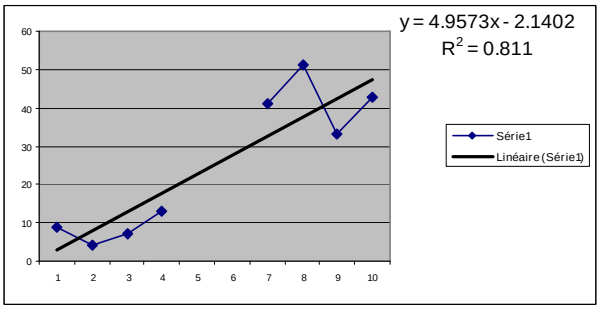
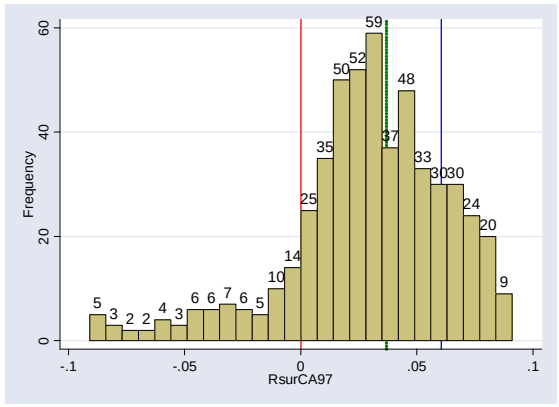
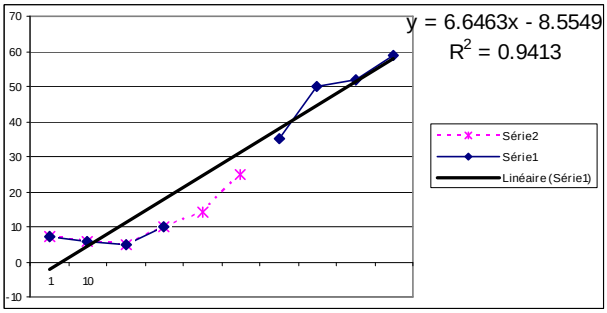
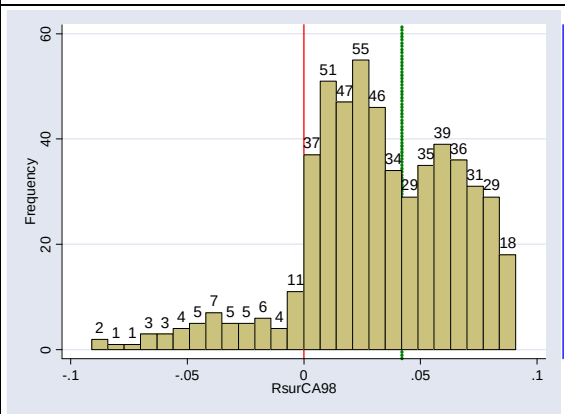
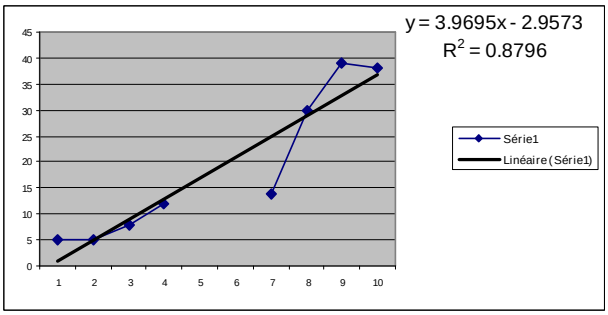
stats	IBsulBCA92	IBsulBCA93	IBsulBCA94	IBsulBCA95	IBsulBCA96	IBsulBCA97	IBsulBCA98	IBsulBCA99	IBsulBCA00	IBsulBCA01	IBsulBCA02	IBsulBCA03	IBsulBCA04
N	288	291	373	418	612	687	736	772	772	673	688	649	597
max	2.053957	2.795608	19.50618	3.048578	18.80364	27.00927	19.82002	33.38549	49.61606	175.3555	29.85712	25.6744	20.05997
min	-3.018	-3.068367	-12.64459	-153.6414	-96.27054	-35.58554	-31.76497	-36.33228	-12.9423	-33.25301	-20.6474	-165.2615	-55.06562
médiane	0.024686	0.0217195	0.0273265	0.028446	0.0381032	0.0369982	0.0421746	0.0451503	0.037465	0.0273167	0.0247104	0.0239509	0.0317812
écart type	0.3082519	0.3740858	1.268191	7.526911	4.153843	1.816903	1.668739	2.240546	2.088142	7.574874	1.740028	6.943544	2.604752
moyenne	0.0717975	0.0479218	0.0839082	-0.2847167	-0.0928947	0.0604884	0.1113035	0.0372576	0.0560477	0.2412572	-0.0245616	-0.3292428	-0.0832799

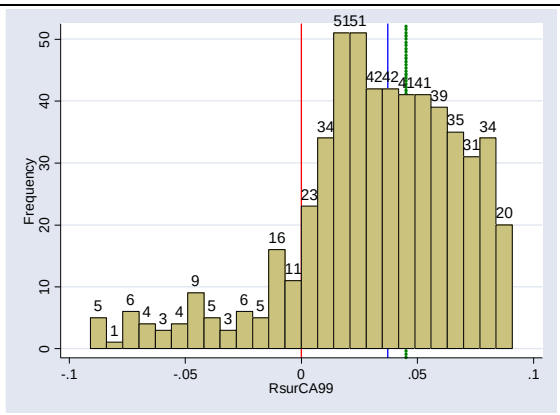
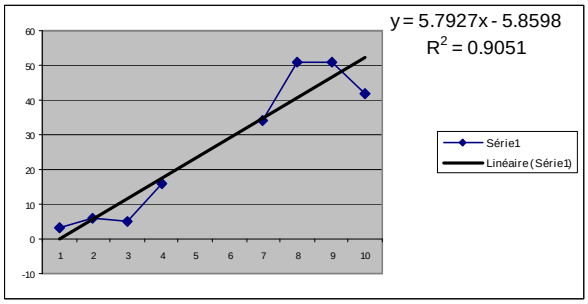
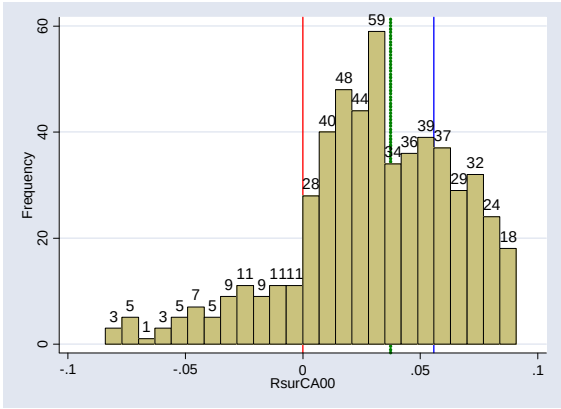
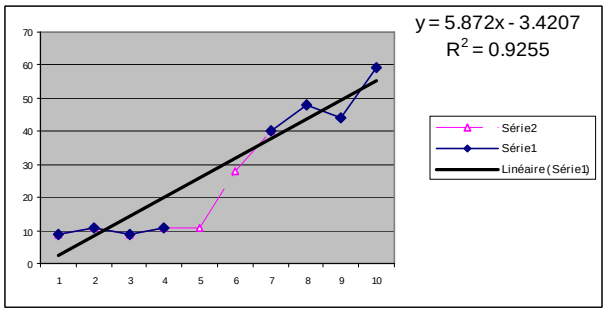
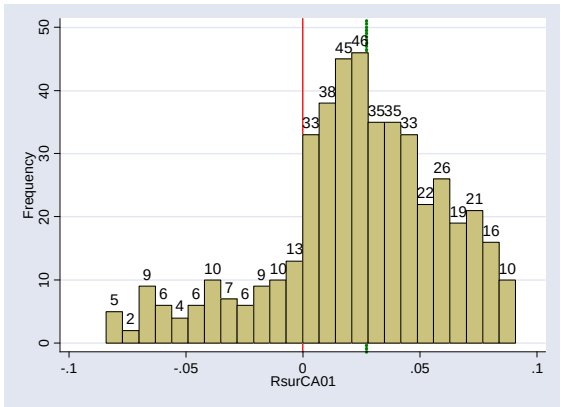
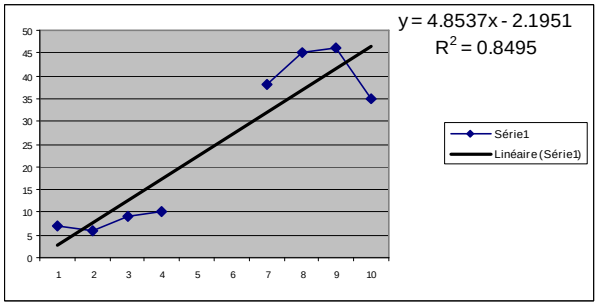
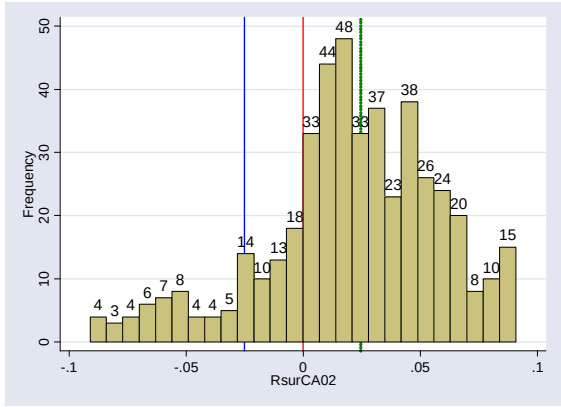
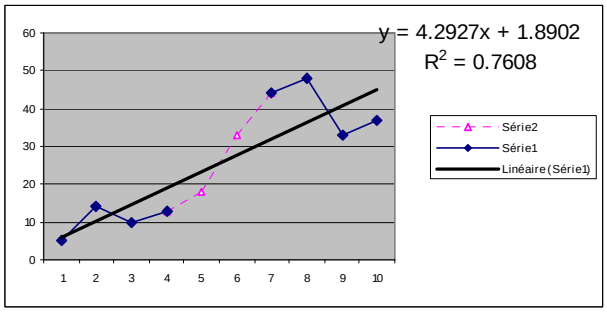
Ensuite, chaque année, une régression linéaire est calculée afin d'estimer par interpolation les effectifs théoriques, et d'en déduire l'irrégularité.

[197] Détail de l'étude menée dans la section VI-C page 226.

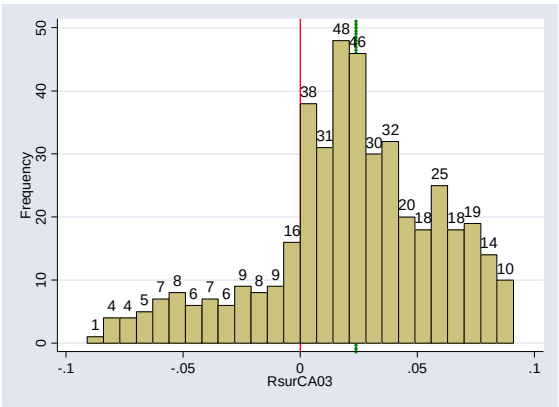
Tableau 127 : Distributions annuelles et effectifs irréguliers de la variable résultat mise à l'échelle par le chiffre d'affaires

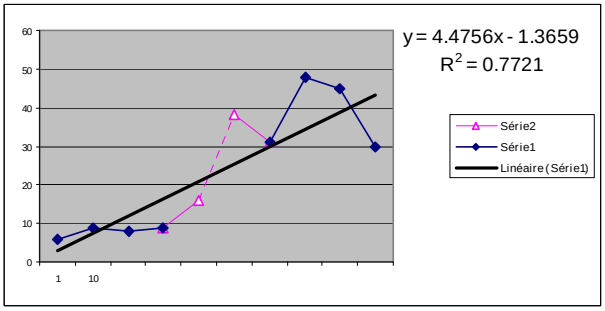
Année	Graphe	interpolation
1992		
	histogram RsurCA92 if RsurCA92<0.091 & RsurCA92>-0.09, width(0.007) start(-0.0910000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.0247, lcolor(blue)) xline(0.0718, lcolor(green) lwidth(thick) lpattern(dot))	
1993		
	histogram RsurCA93 if RsurCA93<0.091 & RsurCA93>-0.09, width(0.007) start(-0.0910000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.0479, lcolor(blue)) xline(0.0217, lcolor(green) lwidth(thick) lpattern(dot))	
1994		

1995		
1996		
1997		
1998		

1999		
2000		
2001		
2002		

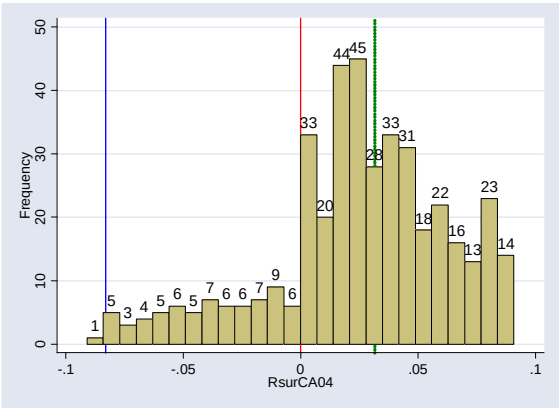
2003

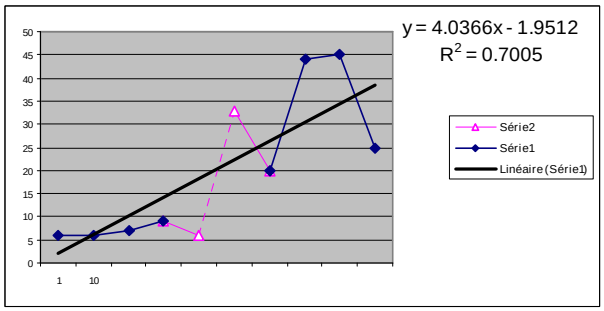




histogram RsurCA03 if RsurCA03<0.091 & RsurCA03>-0.09, width(0.007) start(-0.0910000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.32, lcolor(blue)) xline(0.0239, lcolor(green) lwidth(thick) lpattern(dot))

2004





histogram RsurCA04 if RsurCA04<0.091 & RsurCA04>-0.09, width(0.007) start(-0.0910000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.083, lcolor(blue)) xline(0.0318, lcolor(green) lwidth(thick) lpattern(dot))

Le Tableau 128 récapitule enfin les mesures effectuées à partir des régressions ci-dessus.

Tableau 128 : Calcul des irrégularités de distribution du résultat mis à l'échelle par le chiffre d'affaires

	Intervalle à gauche du seuil				Intervalle à droite du seuil			
	Observation	Estimation	Écart = observation - estimation	irrégularité (en % de l'estimation)	Observation	Estimation	écart	irrégularité (en %)
1992	9	13.22	-4.22	-31,9%	24	15.78	8.22	52.1%
1993	4	12.64	-8.64	-68,4%	21	14.61	6.39	43.7%
1994	7	15.64	-8.64	-55,2%	25	18.61	6.39	34.3%
1995	9	16.78	-7.78	-46,4%	30	20.22	9.78	48.4%
1996	17	22.65	-5.65	-24,9%	25	27.60	-2.60	-9.4%
1997	14	24.68	-10.68	-43,3%	25	31.32	-6.32	-20.2%
1998	7	16.89	-9.89	-58,6%	19	20.86	-1.86	-8.9%
1999	11	23.10	-12.10	-52,4%	23	28.90	-5.90	-20.4%
2000	11	25.94	-14.94	-57,6%	28	31.81	-3.81	-12.0%
2001	13	22.07	-9.07	-41,1%	33	26.93	6.07	22.6%
2002	18	23.35	-5.35	-22,9%	33	27.65	5.35	19.4%
2003	16	21.01	-5.01	-23,9%	38	25.49	12.51	49.1%
2004	6	18.23	-12.23	-67,1%	33	22.27	10.73	48.2%

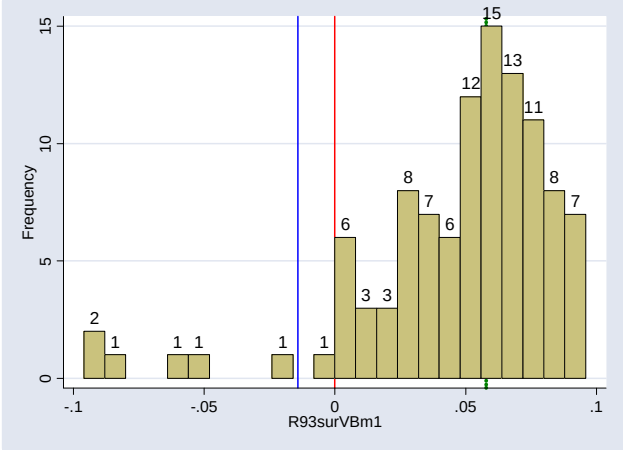
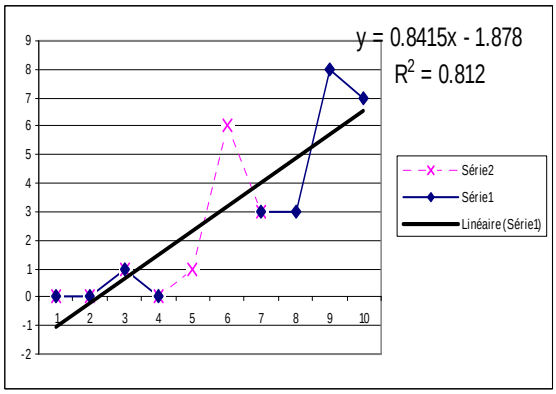
Annexe 10 : Irrégularités mesurées en mettant le résultat à l'échelle par la valeur de marché[198]

Les mesures sont effectuées à partir des distributions annuelles dont les caractéristiques sont récapitulées dans le tableau ci-dessous :

Tableau 129 : Caractéristiques des distributions du résultat mis à l'échelle par la valeur de marché

stats	RsurCA92	RsurCA93	RsurCA94	RsurCA95	RsurCA96	RsurCA97	RsurCA98	RsurCA99	RsurCA00	RsurCA01	RsurCA02	RsurCA03	RsurCA04
N		160	203	291	356	480	494	495	531	650	633	625	582
Min		-2.492	-1.569	-1.161	-3.395	-11.153	-52.740	-47.939	-8.243	-14.647	-2.618	-10.443	-4.266
Max		0.611	0.721	2.567	1.284	4.158	7.250	7.231	14.240	12.724	14.877	15.937	4.124
Moyenne		-0.014	0.021	0.047	0.034	0.046	-0.067	-0.013	0.097	0.033	-0.008	-0.093	0.010
Médiane		0.058	0.049	0.055	0.062	0.070	0.062	0.066	0.055	0.039	0.037	0.050	0.056

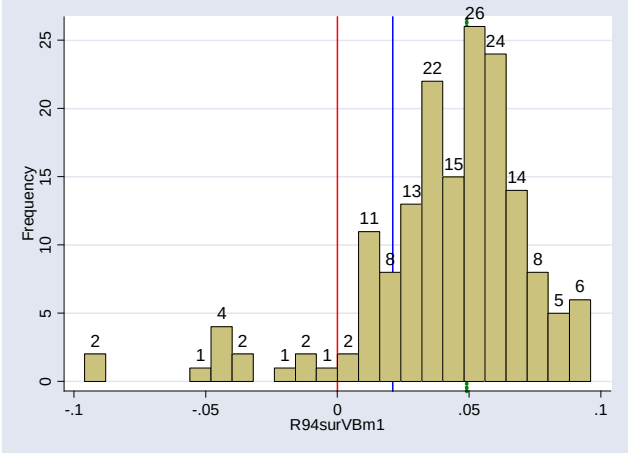
Tableau 130 : Distributions annuelles et effectifs irréguliers de la variable résultat mise à l'échelle par la variable VM

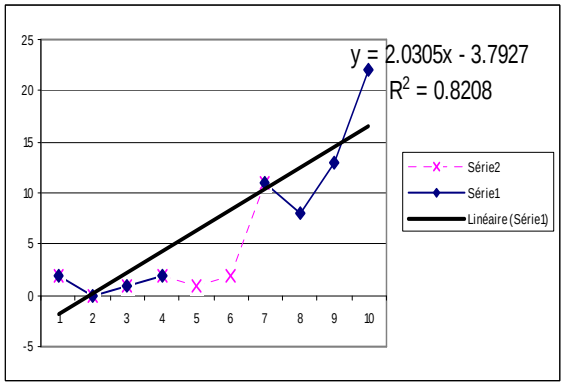
Année	Graphe	Interpolation
1992		
1993		

histogram R93surVBm1 if R93surVBm1<0.096 & R93surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.014, lcolor(blue)) xline(0.058, lcolor(green) lwidth(thick) lpattern(dot))

[198] Détail de l'étude menée dans la section VI-C page 226.

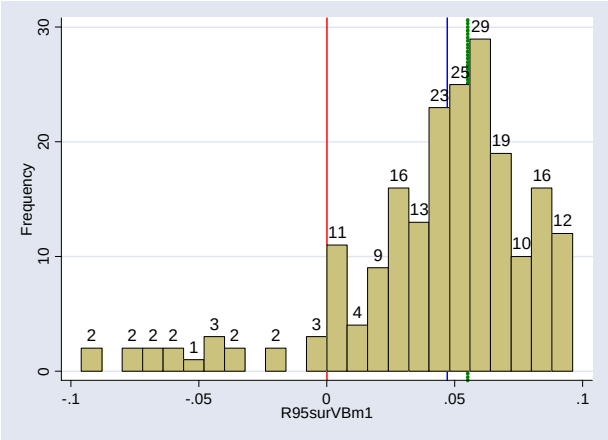
1994

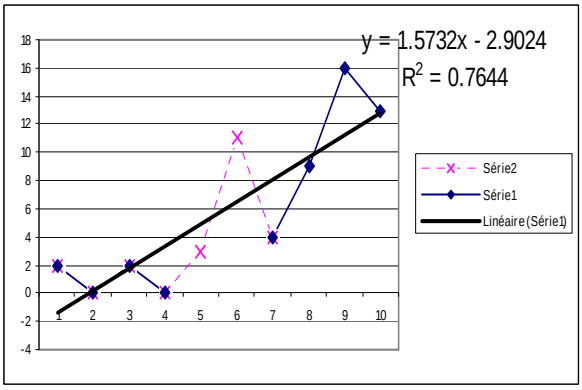




histogram R94surVBm1 if R94surVBm1<0.096 & R94surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.021, lcolor(blue)) xline(0.049, lcolor(green) lwidth(thick) lpattern(dot))

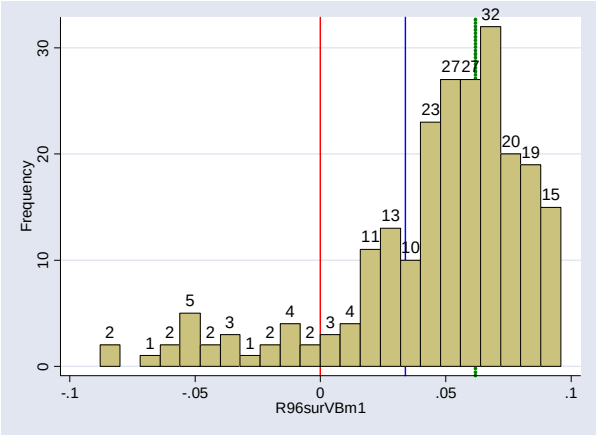
1995

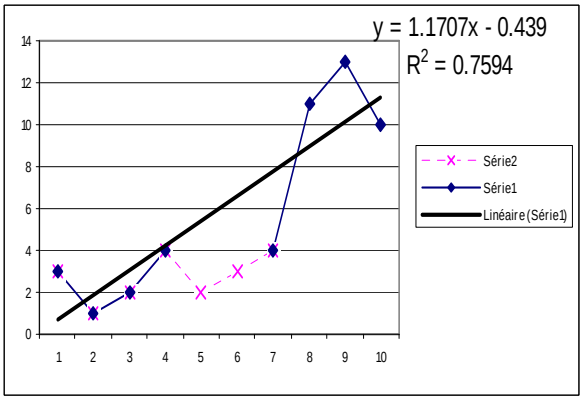




histogram R95surVBm1 if R95surVBm1<0.096 & R95surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.047, lcolor(blue)) xline(0.055, lcolor(green) lwidth(thick) lpattern(dot))

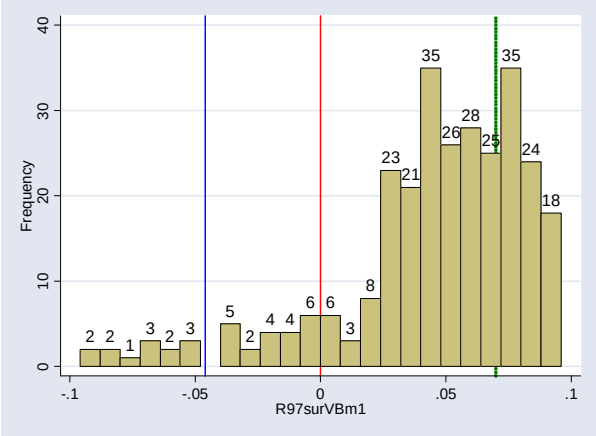
1996

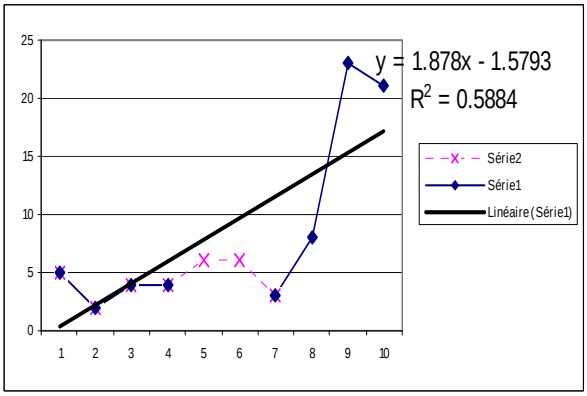




histogram R96surVBm1 if R96surVBm1<0.096 & R96surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.034, lcolor(blue)) xline(0.062, lcolor(green) lwidth(thick) lpattern(dot))

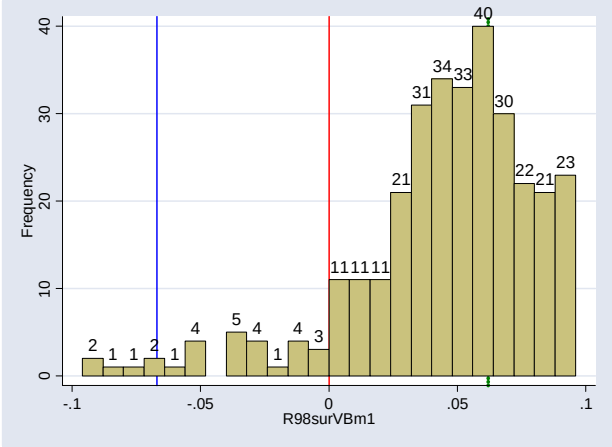
1997

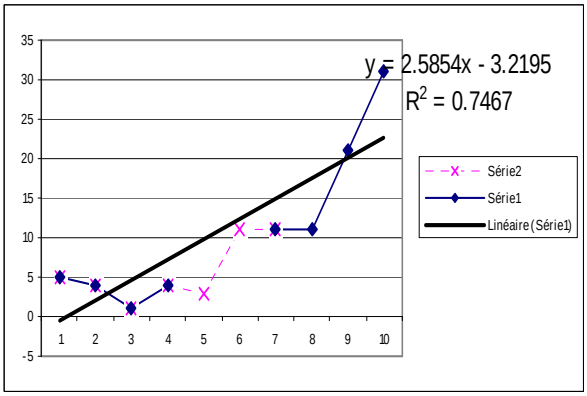




histogram R97surVBm1 if R97surVBm1<0.096 & R97surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.046, lcolor(blue)) xline(0.070, lcolor(green) lwidth(thick) lpattern(dot))

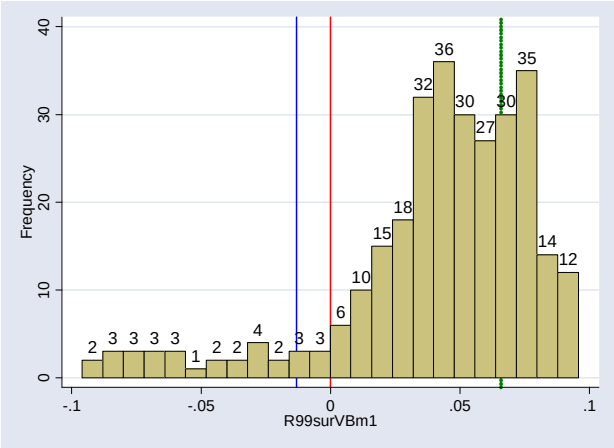
1998

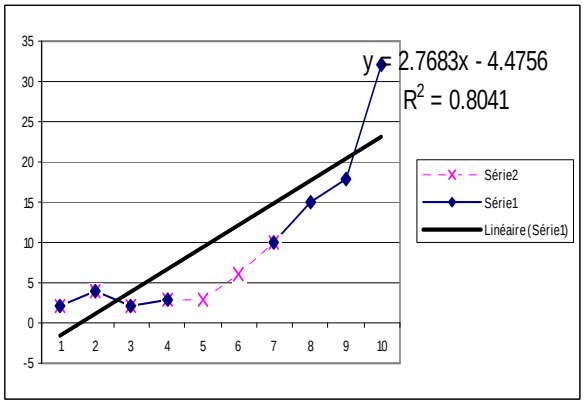




histogram R98surVBm1 if R98surVBm1<0.096 & R98surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.067, lcolor(blue)) xline(0.062, lcolor(green) lwidth(thick) lpattern(dot))

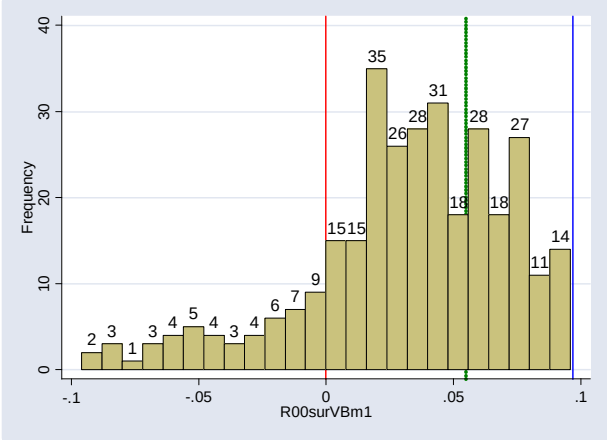
1999

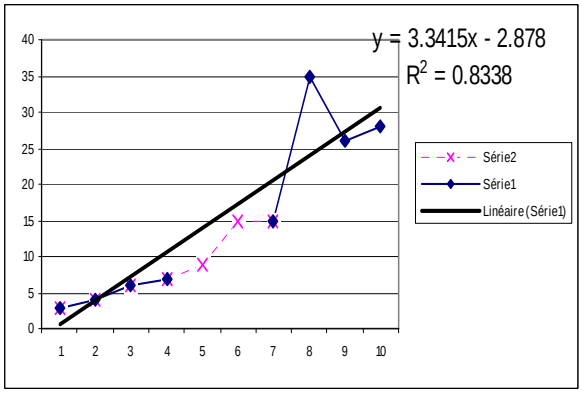




histogram R99surVBm1 if R99surVBm1<0.096 & R99surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.013, lcolor(blue)) xline(0.066, lcolor(green) lwidth(thick) lpattern(dot))

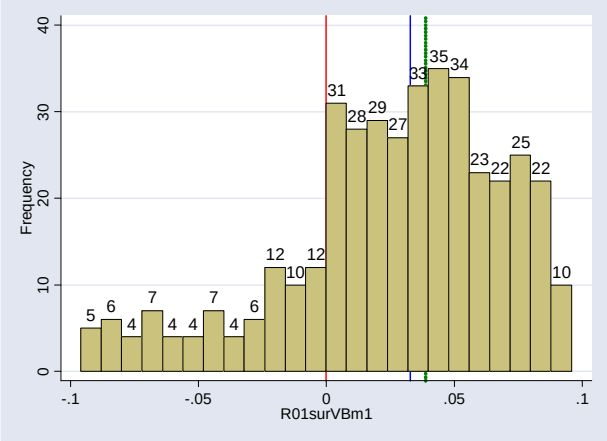
2000

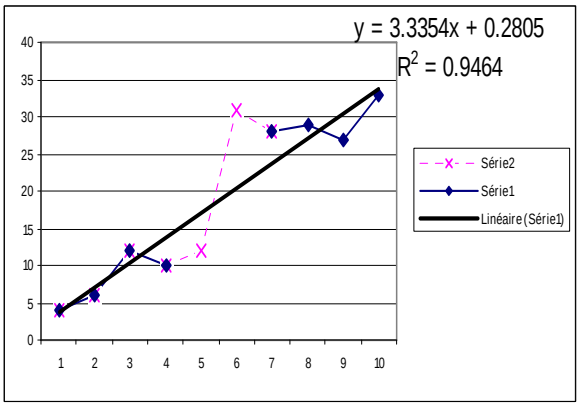




histogram R00surVBm1 if R00surVBm1<0.096 & R00surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.097, lcolor(blue)) xline(0.055, lcolor(green) lwidth(thick) lpattern(dot))

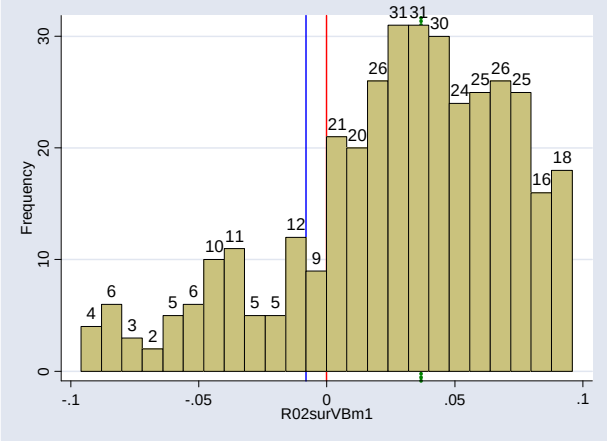
2001

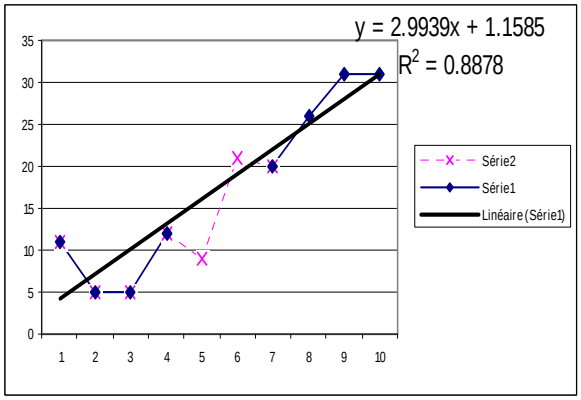




histogram R01surVBm1 if R01surVBm1<0.096 & R01surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.033, lcolor(blue)) xline(0.039, lcolor(green) lwidth(thick) lpattern(dot))

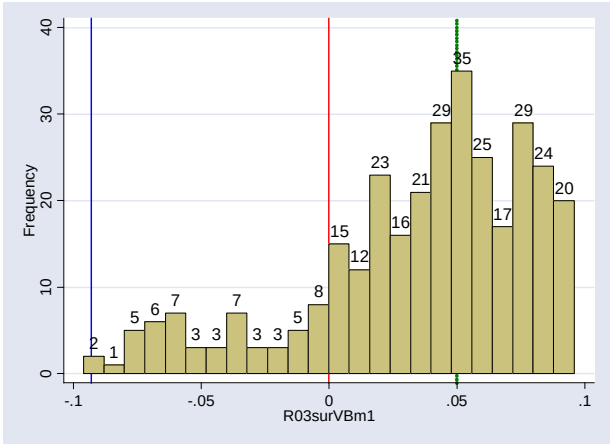
2002

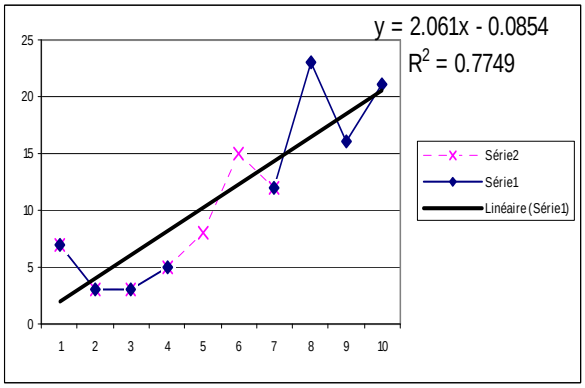




histogram R02surVBm1 if R02surVBm1<0.096 & R02surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.008, lcolor(blue)) xline(0.037, lcolor(green) lwidth(thick) lpattern(dot))

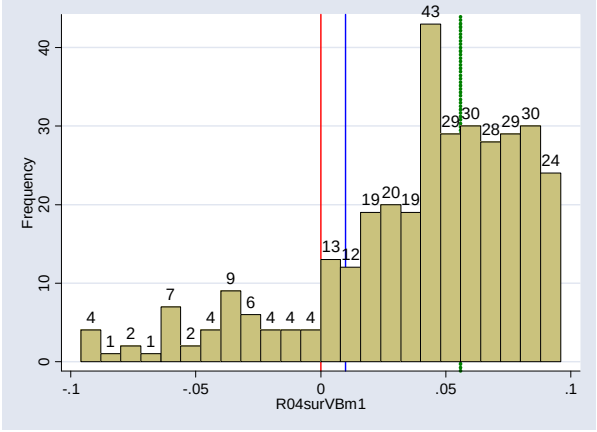
2003

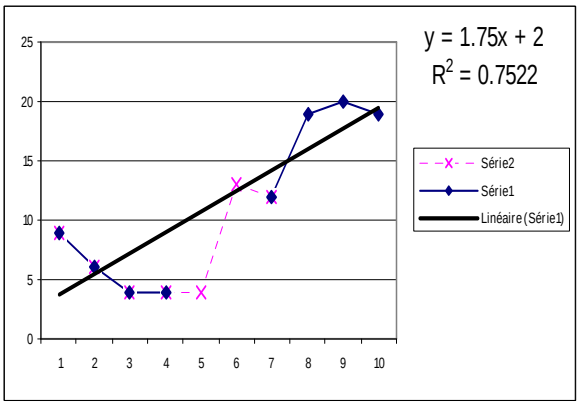




histogram R03surVBm1 if R03surVBm1<0.096 & R03surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(-0.093, lcolor(blue)) xline(0.050, lcolor(green) lwidth(thick) lpattern(dot))

2004





histogram R04surVBm1 if R04surVBm1<0.096 & R04surVBm1>-0.096, width(0.008) start(-0.09600000001) frequency blcolor(black) blwidth(vvthin) addlabels addlabopts(mlabsize(medium)) xline(0, lcolor(red)) xline(0.010, lcolor(blue)) xline(0.056, lcolor(green) lwidth(thick) lpattern(dot))

Le Tableau 131 récapitule enfin les mesures effectuées à partir des régressions ci-dessus.

Tableau 131 : Calcul des irrégularités (résultat mis à l'échelle par la valeur de marché)

	Intervalle à gauche du seuil				Intervalle à droite du seuil			
	Observation	Estimation	écart	irrégularité (en %)	Observation	Estimation	écart	irrégularité (en %)
1992	1	2.33	-1.33	-57.1%	6	3.17	2.83	89.2%
1993	1	6.36	-5.36	-84.3%	2	8.39	-6.39	-76.2%
1994	3	4.96	-1.96	-39.6%	11	6.54	4.46	68.3%
1995	2	5.41	-3.41	-63.1%	3	6.59	-3.59	-54.4%
1996	6	7.81	-1.81	-23.2%	6	9.69	-3.69	-38.1%
1997	3	9.71	-6.71	-69.1%	11	12.29	-1.29	-10.5%
1998	3	9.37	-6.37	-68.0%	6	12.13	-6.13	-50.6%
1999	9	13.83	-4.83	-34.9%	15	17.17	-2.17	-12.6%
2000	12	16.96	-4.96	-29.2%	31	20.29	10.71	52.8%
2001	9	16.13	-7.13	-44.2%	21	19.12	1.88	9.8%
2002	8	10.22	-2.22	-21.7%	15	12.28	2.72	22.1%
2003	4	9.50	-5.50	-57.9%	13	11.25	1.75	15.5%
2004	1	2.33	-1.33	-57.1%	6	3.17	2.83	89.2%

Ces irrégularités sont représentées graphiquement page228.

Annexe 11 : Second examen des valeurs extrêmes au cas par cas[199]

Les entreprises dont le résultat annuel mis à l'échelle par l'actif total (plus précisément par AT-IB) est hors de l'intervalle]-0,5 ; 1[ont été considérées comme des cas extrêmes. Elles font l'objet d'un examen au cas par cas. Cet examen doit se faire distribution par distribution puisque chaque année, le tirage est considéré comme indépendant. L'allure est inchangée, mais les paramètres ponctuels que sont la moyenne et l'écart type dépendent de facteurs conjoncturels.

Année 1995

Tableau 132 : Valeur extrême de l'année 1995

EUROTUNNEL SA	-0.4214
Aucune valeur < -0,5	

En 1995, seule une valeur extrême est à signaler (qui demeure cependant dans les limites]-0,5 ; 1[) : Eurotunnel publie un résultat négatif correspondant à 73 % de son actif de décembre. Cette valeur, certes très importante, n'est pas irrégulière, ni de nature à fausser les paramètres de la distribution.

Année 1996

Tableau 133 : Valeurs extrêmes de l'année 1996

EUROTUNNEL SA	-0.4361
INFOSOURCES	-0.4924
Aucune valeur < -0,5	

En 1996, deux valeurs extrêmes sont à signaler (qui demeurent cependant dans les limites]-0,5 ; 1[) : Eurotunnel et Infosources publient des résultats négatif correspondant à 77 % à 97 % de leurs actifs respectifs. Ces valeurs sont très importantes. Elles ne sont ni irrégulières, ni de nature à fausser les paramètres de la distribution.

[199] Cette annexe détaille les travaux décrits dans la section VIII-A, page 264.

Pour mémoire, Infosources entre dans la base en 1996, et publie des fortes pertes durant 6 ans avant de disparaître de la base. En 2000, l'opérateur de télécommunication belge Belgacom et le groupe Infosources (ex-Infonie) annoncent officiellement la fusion de leurs activités de services d'accès Internet. La nouvelle entité est détenue à hauteur de 75 % par Belgacom et à 25 % par les actionnaires d'Infosources. L'actif reste chaque année important du fait d'augmentations de capital. C'est une valeur atypique, mais pas nécessairement aberrante. Cette société fait partie des valeurs extrêmes plusieurs années.

Année 1997

Tableau 134 : Valeurs extrêmes de l'année 1997

INFOSOURCES	-0.5883
EUROTUNNEL SA	-0.7097

2 valeurs < -0,5

En 1997, pour la seconde année consécutive, ce sont les deux mêmes sociétés qui génèrent des résultats extrêmement négatifs. Pour Eurotunnel, la perte est de 435 alors que l'actif passe de 606 (fin 1996) à 178. Ces valeurs extrêmes ne semblent pas pour autant aberrantes.

Année 1998

Tableau 135 : Valeurs extrêmes de l'année 1998

ILOG SA	-0.3907
INFOSOURCES	-0.4684
INTERACTIF DELTA PRODUCTION	-0.5816

Une seule valeur < -0,5

La société Interactif Delta Production génère une perte égale à 140 %[200] du total actif ! L'actif de 1997 était de 7,2 et il passe à 3,2 après une perte de 4,5. C'est une contre performance extrême mais qui se justifie. La donnée est conservée.

[200] 4,5/3,2 = 1,4

Année 1999

Tableau 136 : Valeurs extrêmes de l'année 1999

Valeurs extrêmes négatives	1999
SELCODIS SA	-0.3366
CYBERSEARCH	-0.3428
ALPHA MOS SA	-0.3479
INTERNATIONAL COMPUTER SA	-0.3541
PHONE SYSTEMS & NETWORK	-0.4510
INFOVISTA SA	-0.4518
INTERACTIF DELTA PRODUCTION	-0.4654
Aucune valeur < -0,5	

graph box ibatiby99, medtype(line)

Valeurs extrêmes positives	1999
SYSTRAN SA	**6.8174**
AZEO	0.5495
FALA SA	0.4630
INITIATIVE ET FINANCE INV SA	0.4377
Une seule valeur > 1	

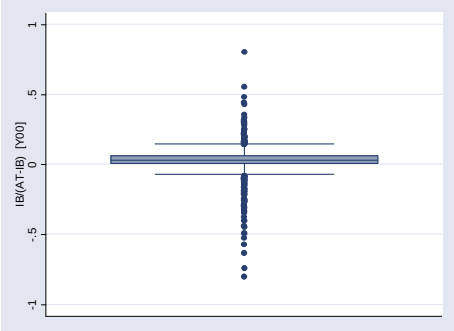
Le résultat de Systran de 1999 avait déjà été identifié comme étant extrême (dans la partie 2). Ce résultat est dû à une réévaluation exceptionnelle, d'une ampleur telle (presque 7 fois le total actif de début d'année) qu'elle fausse les paramètres de la distribution celle année-là. Cette valeur est supprimée de la distribution car elle est manifestement une valeur extrême totalement isolée.

Cette même année, les valeurs extrêmes négatives ne correspondent pas à des cas isolés. Elles ne sont pas supprimées.

Année 2000

Tableau 137 : Valeurs extrêmes de l'année 2000

Valeurs extrêmes négatives	
INTERACTIF DELTA PRODUCTION	-0.3349
CONSORS France	-0.3444
ALPHA MOS SA	-0.3754
IXO	-0.4002
VIKING	-0.4040
BOURSE DIRECT	-0.4367
CAC SYSTEMES SA	-0.4470
PHONE SYSTEMS & NETWORK	-0.4894
INTERCALL SA	-0.4951
INFOSOURCES	**-0.5261**
KALISTO ENTERTAINMENT SA	**-0.5714**
INTEGRA SA	**-0.6315**
WESTERN TELECOM SA	**-0.7410**
EURO RESSOURCES SA	**-0.8006**
5 cas < -0,5	



graph box ibatiby00, medtype(line)

Valeurs extrêmes positives
Aucune valeur > 1

L'année 2000 marque le début d'une crise économique qui va durer plusieurs années. Les performances des entreprises se dégradent très fortement et de manière générale. Il n'est donc pas étonnant de retrouver un plus grand nombre d'entreprises à publier des résultats fortement négatifs. Ces valeurs, même fortement négatives, ne sont pas isolées et n'apparaissent donc pas aberrantes statistiquement.

Année 2001

Tableau 138 : Valeurs extrêmes de l'année 2001

Valeurs extrêmes négatives	
DALET SA	-0.3348
D INTERACTIVE	-0.3370
ACTEOS	-0.3384
GECI INTERNATIONAL	-0.3503
PERFECT TECHNOLOGIES SA	-0.3548
CYBERSEARCH	-0.3568
CAST	-0.3616
CYBERDECK	-0.3646
BD MULTIMEDIA	-0.3731
CAC SYSTEMES SA	-0.3769
NETGEM SA	-0.3774
HUBWOO.COM	-0.3784
BOURSE DIRECT	-0.3900
GENSET SA	-0.4130
PHONE SYSTEMS & NETWORK	-0.4174
HIGHWAVE OPTICAL TECH	-0.4236
NETVALUE SA	-0.4344
AB SOFT SA	-0.4419
WESTERN TELECOM SA	-0.4834
FI SYSTEM SA	-0.4939
INFOSOURCES	**-0.5035**
INTEGRA SA	**-0.5043**
AES CHEMUNEX	**-0.5609**
LIBERTY SURF GROUP SA	**-0.5913**
HI-MEDIA	**-0.6017**
INTERACTIF DELTA PRODUCTION	**-0.6464**
GRANDE PAROISSE	**-0.7181**
GAMELOFT	**-0.7983**
8 cas < -0,5	

graph box ibatiby01, medtype(line)

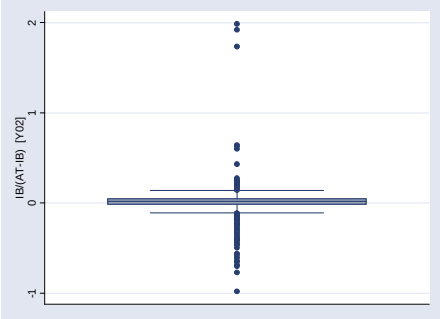
Valeurs extrêmes positives
Aucune valeur > 1

Il apparaît en 2001 une accentuation de la tendance constatée en 2000. Aucune valeur extrême n'est véritablement isolée. Aucune donnée n'est retirée.

Année 2002

Tableau 139 : Valeurs extrêmes de l'année 2002

Valeurs extrêmes négatives	
COM 1 SA	-0.3309
BOURSE DIRECT	-0.3372
ESR	-0.3382
SQLI	-0.3524
ABEL GUILLEMOT	-0.3575
ACCESS COMMERCE	-0.3652
DEVOTEAM SA	-0.3910
BULL SA	-0.3920
SELCODIS SA	-0.4076
NETGEM SA	-0.4136
HUBWOO.COM	-0.4165
HI-MEDIA	-0.4469
CYBER PRESS PUBLISHING	-0.4509
EGIDE SA	-0.4583
PROLOGUE SOFTWARE SA	-0.4611
HIGHWAVE OPTICAL TECH	-0.4976
GENSET SA	**-0.5624**
ARTPRICE.COM	**-0.5783**
BCI NAVIGATION	**-0.6071**
CAC SYSTEMES SA	**-0.6431**
CYBERSEARCH	**-0.6517**
FI SYSTEM SA	**-0.6870**
GAMELOFT	**-0.7006**
CGBI SA	**-0.7725**
AUREA	**-0.9809**
9 cas< -0,5	



graph box ibatiby02, medtype(line)

Valeurs extrêmes positives	
FONCIERE MASSENA	**1.9837**
SOFCO	**1.9193**
PHONE SYSTEMS & NETWORK	**1.7335**
INTERNATIONAL COMPUTER SA	0.6431
AB SOFT SA	0.6302
ADC SIIC	0.5971
FALA SA	0.4298
3 cas > 1	

En 2002, trois sociétés publient des résultats étonnamment élevés et positifs. Ce sont trois valeurs totalement isolées du reste de la population. Cet isolement les rend aberrantes en termes de distribution. Après contrôle, il apparaît que ces trois valeurs sont aberrantes parce que le dénominateur n'est pas un bon indicateur de taille.

Foncière Masséna est une société spécialisée dans la détention et la gestion d'actifs immobiliers d'entreprise. Le patrimoine de la société comprend principalement des actifs immobiliers à usage de locaux commerciaux. Entre 2002 et 2003, le résultat net est très largement supérieur au chiffre d'affaires. Plus précisément, en 2001, le fonds d'investissements immobiliers Partenaires Résidentiels, créé par le Groupe Masséna a acquis auprès de EPP Alpha un bloc de contrôle représentant 87,73 % du capital de

Socim. À l'issue d'une opération de garantie de cours, Partenaires Résidentiels a porté son niveau de détention à 98,08 % du capital. L'arbitrage de la totalité des actifs immobiliers encore détenus par la Socim a été finalisé avec succès au cours du premier trimestre 2004, et la société a distribué l'essentiel du produit des ventes réalisées sous la forme de dividendes exceptionnels. Les indicateurs de taille utilisés pour réduire le résultat (total actif comme chiffre d'affaires) ne sont pas pertinents. La valeur est supprimée de la base de 2002 à 2003.

Tableau 140 : Valeurs extrêmes de l'année 2002 ; Société SOFCO

	IB 1999	IB 2000	IB 2001	IB 2002	IB 2003	IB 2004
SOFCO	0.595315212	0.397897002	0.162	3.58899975	-0.18599999	-0.404999971

À l'origine, Sofco était une société qui exerce une double activité. D'une part, la vente de voitures, neuves ou d'occasion, fabriquées par divers constructeurs, aussi bien français qu'étrangers, et d'autre part, la gestion de biens immobiliers : achat d'espace, construction et exploitation de maisons de repos (pour personnes âgées ou handicapées), d'hôtels et de bâtiments industriels. Cependant, après quelques soucis financiers et judiciaires, elle doit abandonner en 1998 son activité automobile, et se défait peu à peu de ses biens immobiliers. En 2002, le résultat de la société bondit de près de 2100 % ! Cette valeur extrême est une exception dans l'historique de la société, alors que le montant des actifs diminue très sensiblement entre 2001 et 2003 et que le chiffre d'affaires chute brutalement entre 2001 et 2002. En 2002, la société OPEL est condamnée à verser une indemnité de 52M€ à la société SOFCO pour abus de refus d'agrément. Cela n'explique que partiellement la performance exceptionnelle cette année-là. Aucune information plus précise n'a été trouvée sur Internet. En tout état de cause, la mise à l'échelle par le chiffre d'affaires comme par le total actif ne semblent pas pertinents dans cette situation. La société est supprimée de la base en 2002 uniquement.

En ce qui concerne Phone Systems & Network, des recherches sur Internet révèlent que la base Global est incomplète à partir de 2003.

Tableau 141 : Valeurs extrêmes de l'année 2002 ; société Phone Systemes & Network

	2000	2001	2002	2003	2004	2005
IB	-5,748	-2,325	**1,106**	0,378	0,639	-0,725
AT	5,998	3,245	**1,744**	2,532	3,782	4,169

Il apparaît que la performance très importante de 2002 au regard du montant des actifs est exceptionnelle. D'autant plus qu'en 2002, le résultat d'exploitation de la société est

fortement négatif (-1,010). À cette époque, un changement de direction et une réorganisation de l'entreprise eut lieu. Il apparaît donc que le résultat de 2002 ne peut être mis à l'échelle par l'actif total (ni même par le chiffre d'affaires) sans fausser la distribution. Cette valeur est retirée de la base en 2002.

En ce qui concerne les valeurs négatives, comme en 2001, les performances fortement négatives sont nombreuses, mais une seule société apparaît comme isolée par rapport à la population. Il s'agit de la société Aurea.

Aurea est une société holding dont les filiales, jusqu'en 2002, exercent l'activité de distribution de produits destinés au marché dentaire. La société est retirée de la cote du 21 décembre 2001 au 30 août 2002. Durant cette période, le commissaire aux comptes refuse de certifier les états financiers.

Tableau 142 : Valeurs extrêmes de l'année 2002 ; société Aurea

	Résultat										Actif Total									
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
AUREA	0.217062776	-0.034569338	-0.07651863	-0.029257663	-0.962870968	-1.881702494	-0.605000019	-1.901100002	-0.048	0.403999984	7.814711904	7.259973023	6.758607113	7.099814485	6.910525735	5.387562264	8.840999603	0.037	0.056000002	27.65899658

Une enquête menée par l'AMF mènera la haute autorité à sanctionner le dirigeant de l'époque pour fausse information. Durant cette période, Aurea revend tous ses actifs, et ne détient plus alors que de la trésorerie. Le résultat de 2002 (une perte de 1,901 M€) est alors à comparer à un actif non représentatif de 0,037M€. Ce ratio ne semble pas fournir une information pertinente et déforme la distribution des résultats réduits. La valeur est retirée en 2002. Il n'est pas utile de supprimer les années 2003 et 2004 car en 2003 le résultat (faible perte) est proportionné à la valeur des actifs, et en 2004, une augmentation de capital permet à la société d'acquérir de nouvelles participations dans des activités de recyclage. La société ne fait plus partie des valeurs extrêmes. Aurea est donc supprimée en 2002 uniquement.

Une autre société attire l'attention : CGBI est une société de services informatiques qui traverse une période difficile entre 2000 et 2003.

Tableau 143 : Valeurs extrêmes de l'année 2002 ; société CGBI

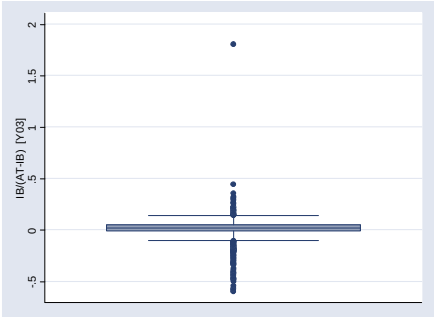
	Résultat								Actif Total							
	1997	1998	1999	2000	2001	2002	2003	2004	1997	1998	1999	2000	2001	2002	2003	2004
CGBI SA	0.556388922	1.142115702	1.483485966	-4.687255802	-8.68200016	**-29.1559906**	**-11.54399967**	2.835000038	5.393531618	9.257940547	19.76653714	51.75032835	47.2869873	**8.585000038**	**8.411999702**	4.401999474

Le résultat est fortement négatif en 2002 et 2003. En 2004, au contraire, la société occupe une valeur extrême positive. Les fluctuations du résultat sont telles, ainsi que les fluctuations du montant des actifs, que l'évolution de cette valeur fausse la distribution. La société est retirée de la base de 2002 à 2004.

Année 2003

Tableau 144 : Valeurs extrêmes de l'année 2003

Valeurs extrêmes négatives	
ARTPRICE.COM	-0.3355
CROSS SYSTEMS CO	-0.3359
MEMSCAP SA	-0.3748
PERFECT TECHNOLOGIES SA	-0.3896
R2I SANTE SA	-0.4055
BCI NAVIGATION	-0.4146
NETGEM SA	-0.4234
EUROTUNNEL SA	-0.4317
HIGHWAVE OPTICAL TECH	-0.4436
AUREA	-0.4615
QUALIFLOW 5	-0.4729
PROLOGUE SOFTWARE SA	-0.4742
COM 1 SA	-0.4801
RISC GROUP	-0.4977
BOURSE DIRECT	-0.5413
TITUS INTERACTIVE SA	-0.5655
CGBI SA	-0.5785
IMECOM GROUP SA	-0.5943
4 cas < -0,5	



graph box ibatiby03, medtype(line)

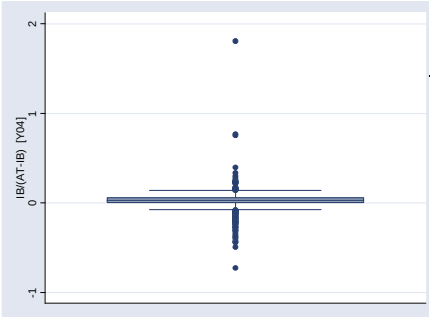
Valeurs extrêmes positives	
FONCIERE MASSENA	**1.8059**
CYBER PRESS PUBLISHING	0.4456
GAMELOFT	0.3562
FALA SA	0.3206
1 cas > 1	

En 2003, la société Foncière Masséna (déjà discutée) apparaît comme un cas isolé de valeur extrême positive. Elle est donc retirée. En ce qui concerne les valeurs négatives, même si les valeurs très négatives sont encore relativement nombreuses, aucune n'est isolée de l'ensemble de la population. Seule CGBI (déjà discutée) est retirée du fait de variations extrêmes.

Année 2004

Tableau 145 : Valeurs extrêmes de l'année 2004

Valeurs extrêmes négatives	
FAROS SA	-0.3021
ARTPRICE.COM	-0.3163
CHAINE ET TRAME SA	-0.3203
CYBER PRESS PUBLISHING	-0.3431
STE FRANCAISE PAPIERS PEINT	-0.3627
RISC GROUP	-0.3809
WAVECOM SA	-0.3817
GENESYS SA	-0.4011
EUROTUNNEL SA	-0.4293
IMECOM GROUP SA	-0.4357
TRANSGENE	-0.4418
R2I SANTE SA	-0.4972
HIGHWAVE OPTICAL TECH	**-0.7301**
1 seul cas < -0,5	



graph box ibatiby04, medtype(line)

Valeurs extrêmes positives	
CGBI SA	**1.8092**
GROUPE PHILIPPE BOSC	0.7653
BULL SA	0.7528
1 seul cas > 1	

En 2004, une seule société (déjà examinée) affiche une valeur extrême positive complètement isolée. Cette valeur est donc éliminée de la base. En ce qui concerne les valeurs négatives, le nombre de valeurs extrêmes diminue et les ordres de grandeur ne sont pas aberrants.

Synthèse

En définitive, 9 données ont été supprimées de la base, correspondant à 5 sociétés. Une société a été supprimée sur trois années du fait de variations extrêmes (2002 à 2004), 6 données (dont une seule négative en 2002) ont été retirées du fait d'une situation isolée de la population, à chaque fois cela s'expliquant par un dénominateur peu significatif. En conclusion, la médiane apparaît comme beaucoup moins sensible aux valeurs extrêmes.

Annexe 12 : Mesures année par année avec une loi Logistique (hypothèse de symétrie)

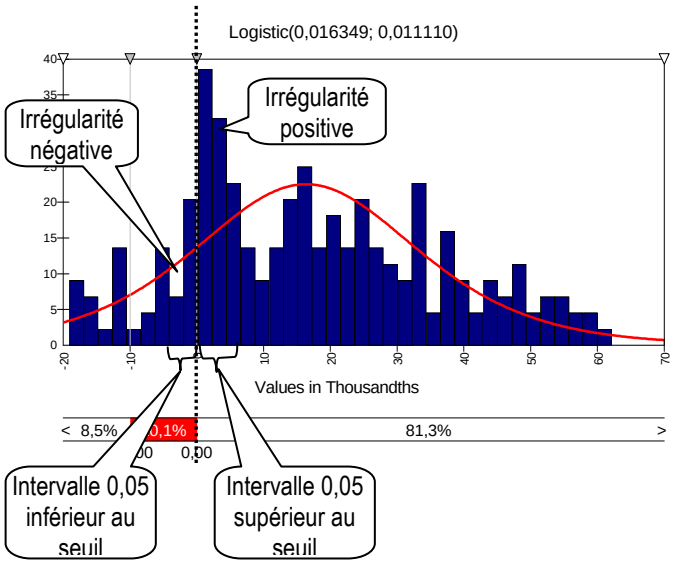
Les mesures des irrégularités faisant référence à la loi Logistique comme allure théorique de la distribution des résultats non manipulés font l'hypothèse que la distribution doit être symétrique. Les mesures présentées ci-dessous détaillent les calculs effectués dans le cadre de la section IX-A (page 304).

Année 1992

Tableau 146 : Ajustement de la distribution 1992 par une loi Logistique

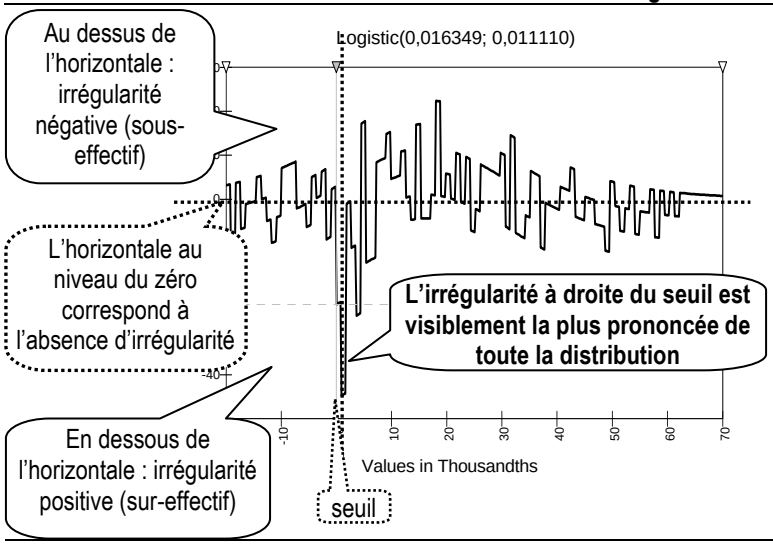
				Détermination des bornes en 1992	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
291	0,028	0,0197	0,0545	-0,0239	0,0633

	Input	Logistic
Function		=RiskLogistic(0,016349 ; 0,011110)
Parameter 1		0,0163492313502607
Parameter 2		0,01110988660224
Minimum	-0,019100	-Infinity
Maximum	0,062100	+Infinity
Mean	0,017218	0,016349
Mode	0,00060000	0,016349
Median	0,015300	0,016349
Std. Deviation	0,018958	0,020151
Variance	0,00035765	0,00040607
Skewness	0,3231	0,0000
Kurtosis	2,3460	4,2000
A-D Test Value		1,686
P Value		< 0,005
Rank		7
K-S Test Value		0,08120
P Value		< 0,01
Rank		6
Chi-Sq Test Value		26,86
P Value		0,0201



Rank 6

Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	13	33
Ajusté	17.67	21.43
Irrégularité	-4,67	11,57
soit	-26.43%	**53.99%**

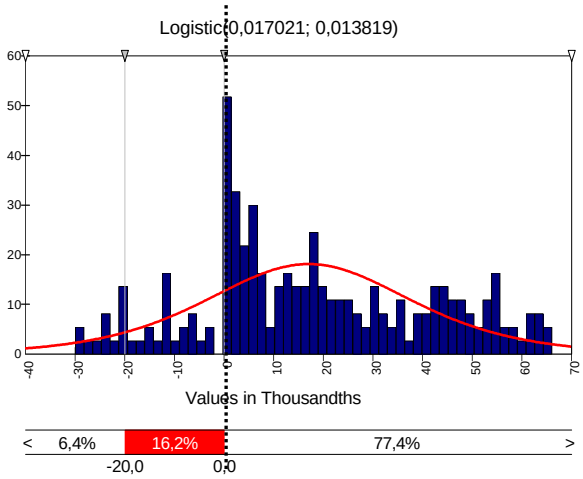
En 1992, l'irrégularité au dessus du résultat nul (54 %) apparaît comme étant la plus importante de toute les irrégularités positives de la distribution. L'irrégularité en dessous du résultat nul est peu prononcée mais négative (-26 %) conformément à la théorie.

Année 1993

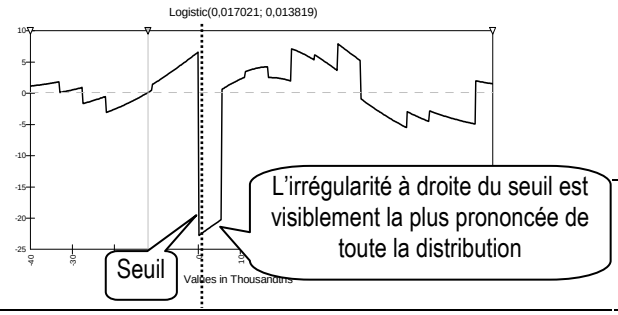
Tableau 147 : Ajustement de la distribution 1993 par une loi Logistique

				Détermination des bornes en 1993	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
296	0,0199	0,0173	0,0616	-0,03198	0,06658

	Input	Logistic
Function		=RiskLogistic(0,017021 ; 0,013819)
Parameter 1		1,70210988045362E-02
Parameter 2		0,013818807197189
Minimum	-0,030000	-Infinity
Maximum	0,066000	+Infinity
Mean	0,017930	0,017021
Mode	0,000000	0,017021
Median	0,015100	0,017021
Std. Deviation	0,023459	0,025065
Variance	0,00054769	0,00062823
Skewness	0,1984	0,0000
Kurtosis	2,2088	4,2000
A-D Test Value		2,370
P Value		< 0,005
Rank		4
K-S Test Value		0,08055
P Value		< 0,01
Rank		3
Chi-Sq Test Value		172,8
P Value		4,7347E-06
Rank		3



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	4	38
Ajusté	11,97	14,57
Irrégularité	-7,97	23,43
soit	**-66,6%**	**160,8%**

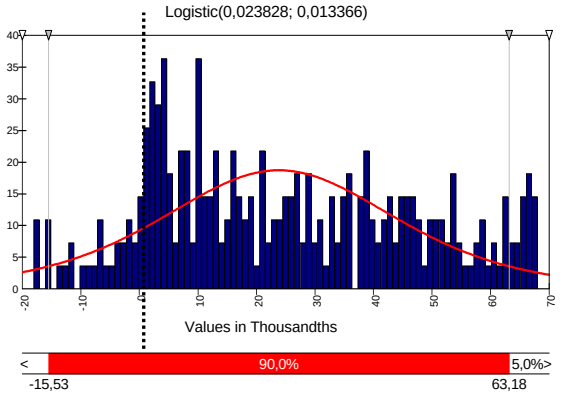
En 1993, l'irrégularité au dessus du résultat nul (161 %) apparaît comme étant la plus importante des irrégularités positives de toute la distribution. L'irrégularité en dessous du résultat nul (-67 %) est également une des plus prononcée parmi les irrégularités négatives.

Année 1994

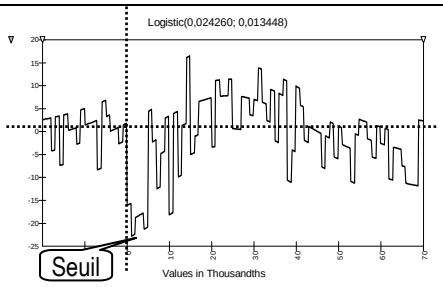
Tableau 148 : Ajustement de la distribution 1994 par une loi Logistique

				Détermination des bornes en 1994	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**Borne sup** (m + 0,8σ)
376	0,029	0,0249	0,0544	-0,01867913	0,068433416

	Input	**Logistic**
Function		=RiskLogistic(0,023828; 0,013366)
Parameter 1		0,023828026
Parameter 2		0,013365558
Minimum	-0,0181	-Infinity
Maximum	0,068	+Infinity
Mean	0,02477	0,023828
Mode	0,0037000	0,023828
Median	0,0215	0,023828
Std. Deviation	0,022337	0,024242
Variance	0,00049713	0,0005877
Skewness	0,2571	0
Kurtosis	2,0278	4,2
A-D Test Value		3,681
P Value		< 0,005
Rank		6
K-S Test Value		0,08375
P Value		< 0,01
Rank		5
Chi-Sq Test Value		52,46
P Value		9,2289E-06
Rank		6



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	10	41
Ajusté	11	14,34
Irrégularité	-1	26,66
soit	**-9,1 %**	**185,9 %**

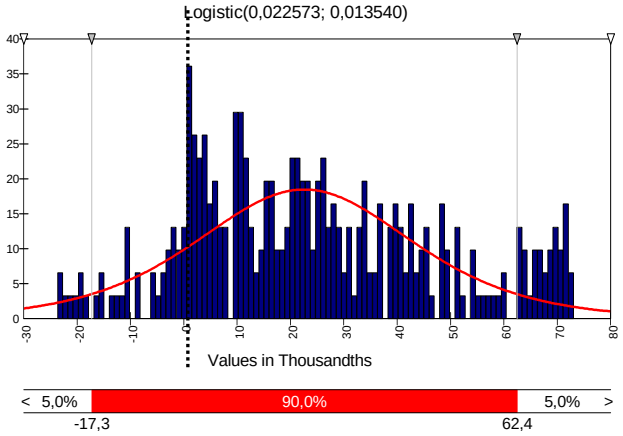
En 1994, l'irrégularité au dessus du résultat nul (186 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul est quasiment inexistante (mais demeure négative).

Année 1995

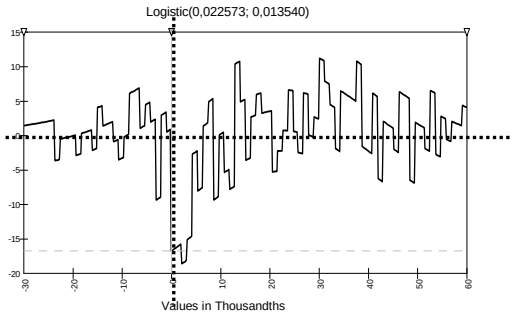
Tableau 149 : Ajustement de la distribution 1995 par une loi Logistique

				Détermination des bornes en 1995	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
422	0,0105996	0,0227025	0,0618258	-0,026758	0,0721632

	Input	Logistic
Function		=RiskLogistic(0,022573; 0,013540)
Parameter 1		0,022573297
Parameter 2		0,013540092
Minimum	-0,0237	-Infinity
Maximum	0,073	+Infinity
Mean	0,023882	0,022573
Mode	0,0014000	0,022573
Median	0,0211	0,022573
Std. Deviation	0,023293	0,024559
Variance	0,00054083	0,00060314
Skewness	0,3741	0
Kurtosis	2,4138	4,2
A-D Test Value		2,745
P Value		< 0,005
Rank		6
K-S Test Value		0,05889
P Value		< 0,01
Rank		6
Chi-Sq Test Value		52,83
P Value		0,00001521
Rank		7



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	4	40
Ajusté	11,97	17,55
Irrégularité	-7,97	22,45
Soit	**-66,5%**	**127,9%**

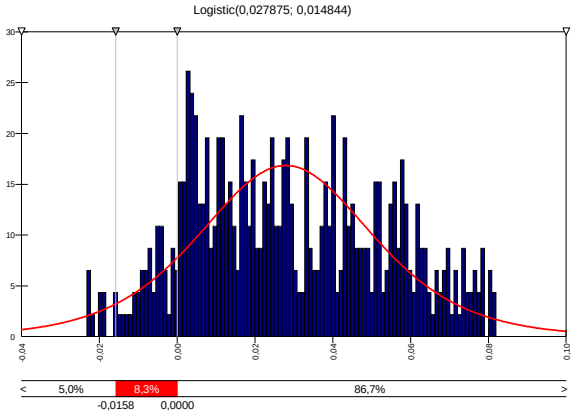
En 1995, l'irrégularité au dessus du résultat nul (128 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul (-67 %) est négative mais moins prononcée.

Année 1996

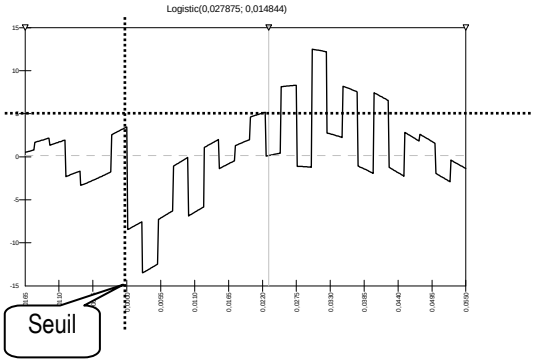
Tableau 150 : Ajustement de la distribution en 1996

				Détermination des bornes en 1996	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
616	0,019611753	0,028484924	0,06652259	-0,024733148	0,081702996

	Input	Logistic
Function		=RiskLogistic(0,027875; 0,014844)
Parameter 1		0,027875056
Parameter 2		0,014843863
Minimum	-0,0233	-Infinity
Maximum	0,0819	+Infinity
Mean	0,028667	0,027875
Mode	0,0026000	0,027875
Median	0,0261	0,027875
Std. Deviation	0,024914	0,026924
Variance	0,00061937	0,00072489
Skewness	0,2014	0
Kurtosis	2,1236	4,2
A-D Test Value		4,186
P Value		< 0,005
Rank		6
K-S Test Value		0,0609
P Value		< 0,01
Rank		5
Chi-Sq Test		
Value		71,22
P Value		1,1497E-07
Rank		6



Distribution des irrégularités



Min	0	-0,005
Max	0,005	0
Observé	45	14
Ajusté	20,43	15,97
Irrégularité	24,57	-1,97
soit	**120,3%**	**-12,3%**

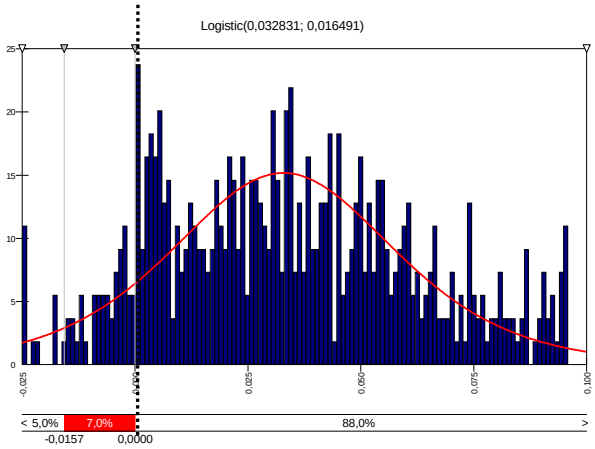
En 1996, l'irrégularité au dessus du résultat nul (120 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul (-12 %) est négative mais très peu prononcée.

Année 1997

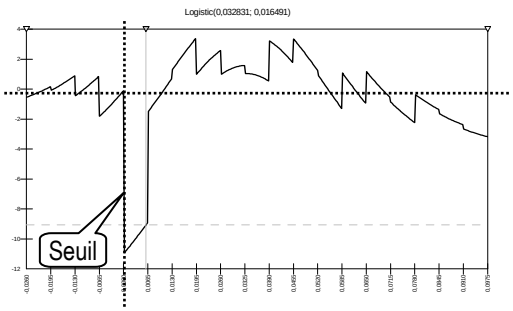
Tableau 151 : Ajustement de la distribution en 1997

				Détermination des bornes en 1997	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
702	0,023705127	0,033275467	0,079087085	-0,029994201	0,096545135

	Input	Logistic
Function		=RiskLogistic(0,032831; 0,016491)
Parameter 1		0,032830946
Parameter 2		0,016491105
Minimum	-0,029	-Infinity
Maximum	0,0958	+Infinity
Mean	0,033726	0,032831
Mode	0,010800	0,032831
Median	0,032	0,032831
Std. Deviation	0,028144	0,029912
Variance	0,0007907	0,0008947
Skewness	0,243	0
Kurtosis	2,3801	4,2
A-D Test Value		2,769
P Value		< 0,005
Rank		6
K-S Test Value		0,04726
P Value		< 0,01
Rank		4
Chi-Sq Test		
Value		65,08
P Value		3,8433E-06
Rank		6



Distribution des irrégularités



Min	0	-0,005
Max	0,005	0
Observé	47	21
Ajusté	20,33	16,16
Irrégularité	26,67	4,84
soit	**131,2%**	**30,0%**

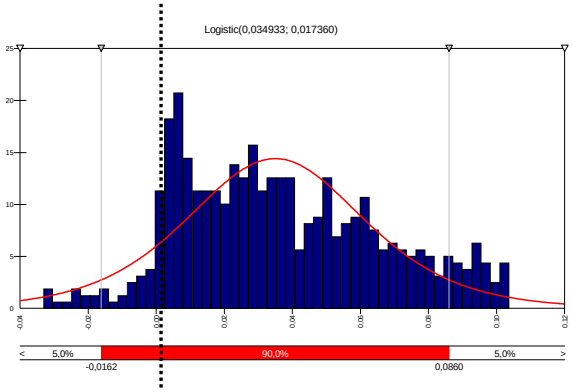
En 1997, l'irrégularité au dessus du seuil (130 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du seuil (positive) est, pour la seule fois de toute la période étudiée, de signe opposé à ce que prédit la théorie.

Année 1998

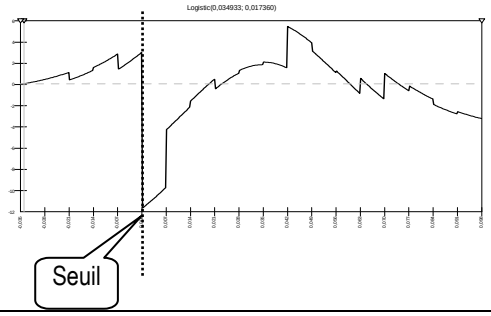
Tableau 152 : Ajustement de la distribution 1998 par une loi Logistique

				Détermination des bornes en 1998	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne inf** (m + 0,8σ)
753	0,030276579	0,034957356	0,086513115	-0,034253135	0,104167848

	Input	Logistic
Function		=RiskLogistic(0,034933; 0,017360)
Parameter 1		0,034932864
Parameter 2		0,017360262
Minimum	-0,0332	-Infinity
Maximum	0,1036	+Infinity
Mean	0,036417	0,034933
Mode	0,0040000	0,034933
Median	0,0324	0,034933
Std. Deviation	0,029652	0,031488
Variance	0,00087774	0,0009915
Skewness	0,3313	0
Kurtosis	2,3962	4,2
A-D Test Value		5,091
P Value		< 0,005
K-S Test Value		0,06173
P Value		< 0,01
Chi-Sq Test Value		105,7
P Value		0



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	9	46
Ajusté	15,58	19,41
Irrégularité	-6,58	26,59
soit	**-42,2%**	**137%**

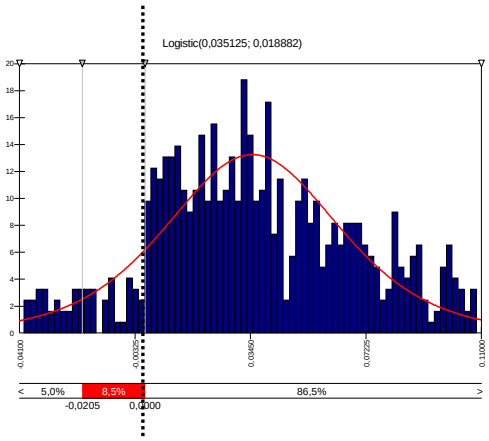
En 1998, l'irrégularité au dessus du résultat nul (137 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul (-42 %) est sensiblement moins prononcée.

Année 1999

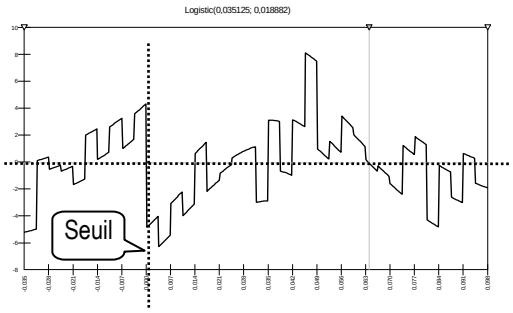
Tableau 153 : Ajustement de la distribution 1999 par une loi Logistique

				Détermination des bornes en 1999	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
797	0,029889593	0,034579113	0,094725248	-0,041201086	0,110359311

	Input	Logistic
Function		=RiskLogLogistic(-0,33276; 0,36719; 19,511)
Parameter 1		-0,332760426
Parameter 2		0,367193211
Parameter 3		19,51076169
Minimum	-0,0396	-0,33276
Maximum	0,1083	+Infinity
Mean	0,03589	0,036024
Mode	0,022400	0,032507
Median	0,03295	0,034433
Std. Deviation	0,032613	0,034463
Variance	0,0010619	0,0011877
Skewness	0,1095	0,4545
Kurtosis	2,5516	4,7171
A-D Test Value		1,766
P Value		N/A
Rank		2
K-S Test Value		0,03783
P Value		N/A
Rank		1
Chi-Sq Test Value		46,66
P Value		0,0025
Rank		1



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	34	8
Ajusté	21,05	17,35
Irrégularité	12,95	-9,35
soit	**61,5%**	**-54,9%**

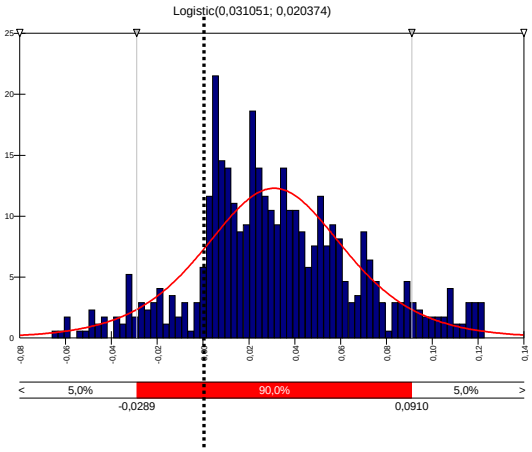
En 1999, l'irrégularité au dessus du résultat nul (62 %) apparaît comme étant la plus importantes de toute la distribution. L'irrégularité en dessous du résultat nul (-55 %) est également très prononcée.

Année 2000

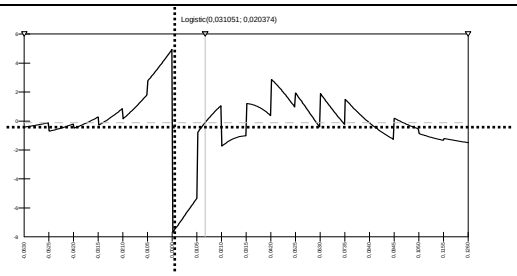
Tableau 154 : Ajustement de la distribution 2000 par une loi Logistique

				Détermination des bornes en 2000	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
791	0,018301021	0,028066391	0,11852293	-0,066751952	0,122884735

	Input	Logistic
Function		=RiskLogLogistic(-0,34814; 0,37849; 18,662)
Parameter 1		-0,348143371
Parameter 2		0,378489277
Parameter 3		18,66209
Minimum	-0,0661	-0,34814
Maximum	0,1227	+Infinity
Mean	0,032188	0,032139
Mode	0,0062000	0,028177
Median	0,0283	0,030346
Std. Deviation	0,036313	0,037172
Variance	0,0013166	0,0013817
Skewness	0,2033	0,476
Kurtosis	3,0958	4,768
A-D Test Value		1,399
P Value		N/A
Rank		1
K-S Test Value		0,05047
P Value		N/A
Rank		1
Chi-Sq Test Value		69,32
P Value		1,5462E-06
Rank		1



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	9	35
Ajusté	21,18	24,79
Irrégularité	-12,18	10,21
soit	**-57,5%**	**41,2%**

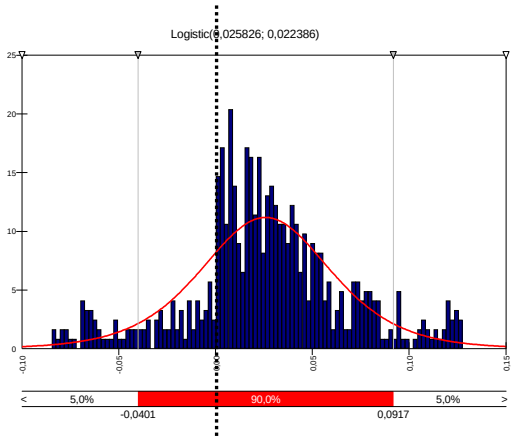
En 2000, les irrégularités au dessus et au dessous du résultat nul apparaissent comme étant les plus importantes de toute la distribution.

Année 2001

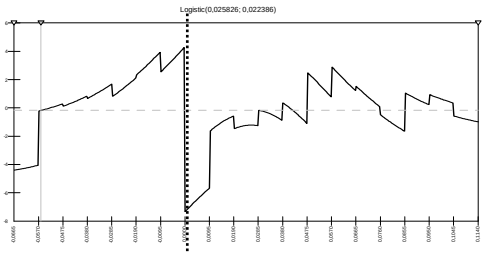
Tableau 155 : Ajustement de la distribution 2001 par une loi Logistique

				Détermination des bornes en 2001	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
734	0,000984002	0,021429188	0,132862575	-0,084860872	0,127719248

	Input	Logistic
Function	N/A	=RiskLogistic(0,025826 ; 0,022386)
Parameter 1	N/A	0,025826398
Parameter 2	N/A	0,022386011
Minimum	-0,0846	-Infinity
Maximum	0,1275	+Infinity
Mean	0,025886	0,025826
Mode	0,021400	0,025826
Median	0,02365	0,025826
Std. Deviation	0,041164	0,040604
Variance	0,0016915	0,0016487
Skewness	-0,0589	0
Kurtosis	3,5192	4,2
A-D Test Value	N/A	2,624
P Value	N/A	< 0,005
Rank	N/A	1
K-S Test Value	N/A	0,07008
P Value	N/A	< 0,01
Rank	N/A	1
Chi-Sq Test Value	N/A	68,94
P Value	N/A	9,6849E-07
Rank	N/A	1



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	13	42
Ajusté	22,16	24,88
Irrégularité	-9,16	17,12
soit	**-41,3%**	**68,8%**

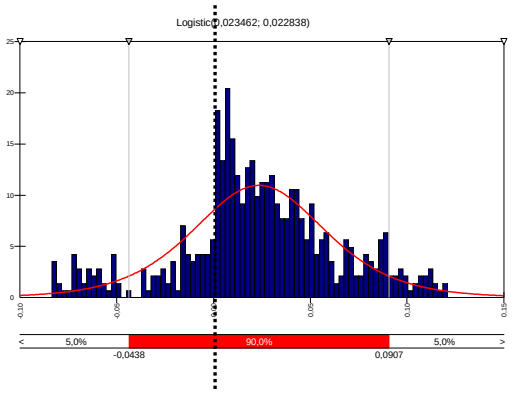
En 2001, l'irrégularité au dessus du résultat nul (69 %) apparaît comme étant la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul (-41 %) est également la plus importante des irrégularités négatives. L'année 2001 est, comme 2000, une année exemplaire. Elle semble coller parfaitement à la théorie. La loi ajuste au mieux la distribution empirique, et les irrégularités à droite et à gauche du résultat nul sont les plus prononcées de la distribution.

Année 2002

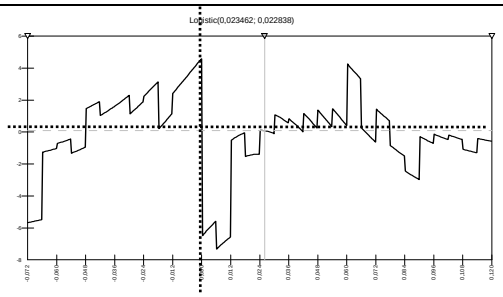
Tableau 156 : Ajustement de la distribution 2002 par une loi Logistique

				Détermination des bornes en 2002	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
693	2,8985	0,0172	0,1305	-0,087258974	0,121577991

	Input	Logistic
Function		=RiskLogistic(0,023462; 0,022838)
Parameter 1		0,023461936
Parameter 2		0,022837834
Minimum	-0,0837	-Infinity
Maximum	0,1209	+Infinity
Mean	0,023061	0,023462
Mode	0,0078000 [est]	0,023462
Median	0,02105	0,023462
Std. Deviation	0,041367	0,041423
Variance	0,0017082	0,0017159
Skewness	-0,1981	0
Kurtosis	3,2556	4,2
A-D Test Value		2,454
P Value		< 0,005
Rank		1
K-S Test Value		0,06751
P Value		< 0,01
Rank		1
Chi-Sq Test		
Value		102,5
P Value		0
Rank		3



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	9	43
Ajusté	22,3	24,73
Irrégularité	-13,3	18,27
soit	**-59,6%**	**73,9%**

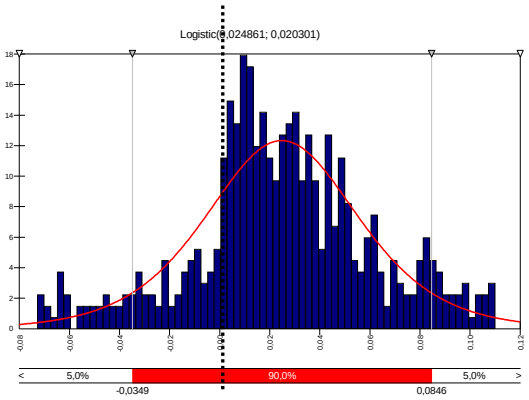
En 2002, les deux irrégularités, au dessus et au dessous du résultat nul, sont les plus importantes de toute la distribution.

Année 2003

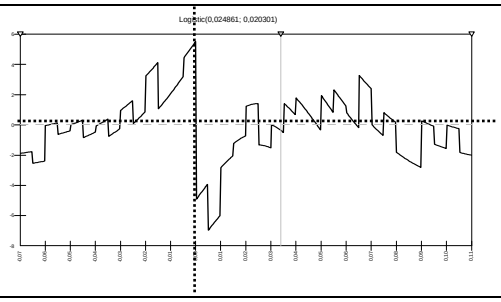
Tableau 157 : Ajustement de la distribution 2003 par une loi Logistique

				Détermination des bornes en 2003	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
657	2,8985	0,0185	0,1161	-0,074456081	0,111362935

	Input	**Logistic**
Function	N/A	=RiskLogistic(0,024861; 0,020301)
Parameter 1	N/A	0,024861157
Parameter 2	N/A	0,020300699
Minimum	-0,0728	-Infinity
Maximum	0,1099	+Infinity
Mean	0,024769	0,024861
Mode	0,014000	0,024861
Median	0,0239	0,024861
Std. Deviation	0,036553	0,036821
Variance	0,0013335	0,0013558
Skewness	-0,1278	0
Kurtosis	3,2129	4,2
A-D Test Value	N/A	1,424
P Value	N/A	< 0,005
Rank	N/A	1
K-S Test Value	N/A	0,05363
P Value	N/A	< 0,01
Rank	N/A	1
Chi-Sq Test Value	N/A	43,18
P Value	N/A	0,003
Rank	N/A	1



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	8	35
Ajusté	20,68	23,65
Irrégularité	-12,68	11,35
soit	**-61,3%**	**48,0%**

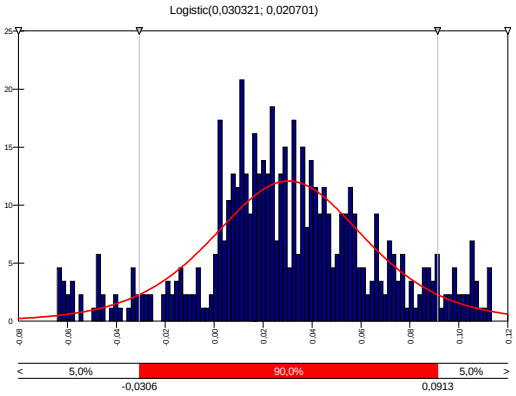
En 2003, les irrégularités au dessus et au dessous du résultat nul sont les plus importantes de toute la distribution. Elles sont conformes à la théorie.

Année 2004

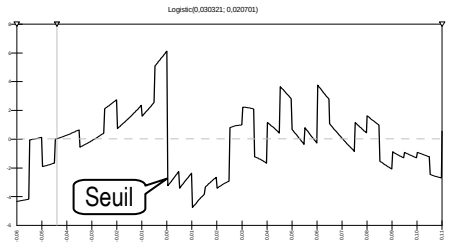
Tableau 158 : Ajustement de la distribution 2004 par une loi Logistique

				Détermination des bornes en 2004	
Nombre	**Moyenne**	**médiane (m)**	**écart type**	**borne inf** (m - 0,8σ)	**borne sup** (m + 0,8σ)
609	0,008966306	0,025613065	0,113881949	-0,065492494	0,116718624

	Input	Logistic
Function		=RiskLogistic(0,030321; 0,020701)
Parameter 1		0,030320823
Parameter 2		0,020700531
Minimum	-0,0643	-Infinity
Maximum	0,1133	+Infinity
Mean	0,03037	0,030321
Mode	0,024500	0,030321
Median	0,0284	0,030321
Std. Deviation	0,037124	0,037547
Variance	0,0013754	0,0014097
Skewness	-0,1085	0
Kurtosis	3,1425	4,2
A-D Test Value	N/A	1,41
P Value	N/A	< 0,005
Rank	N/A	1
K-S Test Value	N/A	0,05483
P Value	N/A	< 0,01
Rank	N/A	1
Chi-Sq Test Value	N/A	33,82
P Value	N/A	0,0379
Rank	N/A	1



Distribution des irrégularités



Min	-0,005	0
Max	0	0,005
Observé	3	26
Ajusté	16,61	19,3
Irrégularité	-13,61	6,7
soit	-81,9%	34,7%

En 2004, les irrégularités au dessus et au dessous du résultat nul apparaissent comme étant les plus importantes de toute la distribution. De signe positif pour l'irrégularité à droite du seuil, et de signe négatif pour l'irrégularité à gauche, elles sont conformes aux prévisions de la théorie.

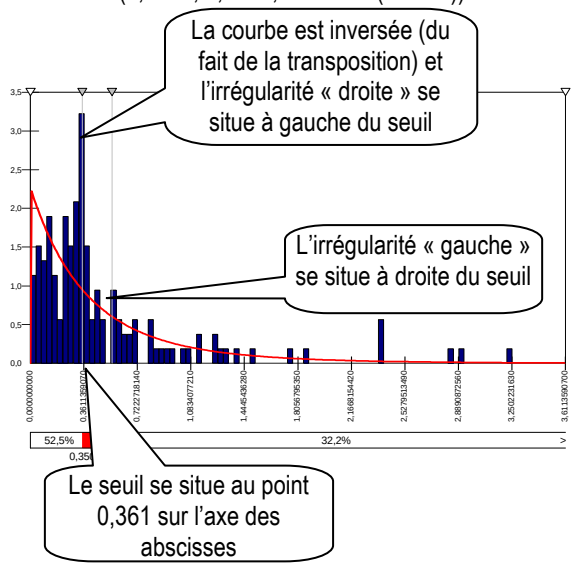
Annexe 13 : Mesures année par année avec une loi Pareto 2 (hypothèse de non symétrie)

Les mesures des irrégularités faisant référence à la loi Pareto 2 comme allure théorique de la distribution des résultats non manipulés s'affranchissent de l'hypothèse de symétrie. Mais les ajustements ne sont faits que sur la demi distribution gauche transposée (multipliée par -1 afin de n'avoir que des valeurs positives). Les mesures présentées ci-dessous détaillent les calculs effectués dans le cadre de la section IX-A (page 308). Les calculs sont présentés année par année.

Année 1992

Tableau 159 : Ajustement de la demi-distribution gauche 1992 par une loi Pareto2

	Distribution de IB/(AT-IB) en 1992 (rappel)	Input = demi-distribution centrée, réduite, et transposée	Pareto2
N	291	146	
Function			=RiskPareto2(2,6586; 6,0210; RiskShift(0.0000))
Shift			1E-10
Parameter 1			2,658574192
Parameter 2			6,021048825
Minimum	-0,2224	0.0000	0.0000
Maximum	0,2136	4,4446	+Infinity
Mean	0,02806271	0,53112	0,52949
Mode		0,11116	0.0000
Median	0,019672663	0,33463	0,32437
Std. Deviation	0,054474404	0,66217	0,64792
Variance		0,43546	0,4198
Skewness		3,0576	3,7985
Kurtosis		14,1826	38,3198
A-D Test Value			+Infinity
P Value			N/A
Rank			1
K-S Test Value			0,1139
P Value			N/A
Rank			1
Chi-Sq Test Value			49,53
P Value			1,6864E-06
Rank			1



Position du résultat nul	**0,361135907**
[(résultat nul – médiane)/écart type] x (-1)	

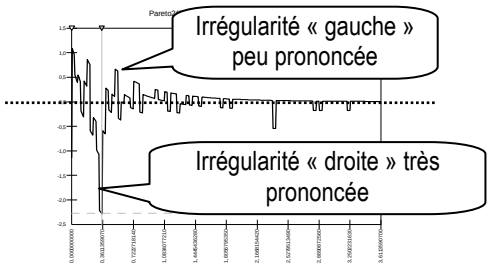
L'irrégularité au seuil du résultat nul est observable graphiquement (lecture au seuil 0,3611 dans la distribution transformée). En revanche, il apparaît aussi que le sommet de la distribution (donc la partie gauche du graphique représentant la demi-distribution)

semble moins bien ajusté par la loi. De manière générale, il apparaît que la loi Pareto2 ajuste moins bien les premières années de l'échantillon.

Autour du seuil du résultat nul, l'irrégularité à droite (donc à gauche du seuil 0,361 sur la distribution ajustée) correspondant aux entreprises faiblement bénéficiaires est très prononcée. En revanche, l'irrégularité à gauche (donc à droite du seuil 0,361 sur la distribution ajustée) correspondant aux entreprises faiblement déficitaires est beaucoup moins prononcée. Il apparaît cependant visuellement que les irrégularités sont conformes à la théorie puisqu'il y a une surreprésentation des entreprises faiblement bénéficiaires et une sous représentation des entreprises faiblement déficitaires.

L'étude de la distribution des erreurs (Tableau 160) permet de confirmer l'analyse graphique et de construire le graphe des irrégularités. Sur ce graphe, il apparaît que l'irrégularité à droite du résultat nul (donc à gauche du trait vertical correspondant au seuil 0,361 de la distribution transformée) est la plus forte de toute la distribution (le « trou » est le plus « profond » sur le graphique). En revanche, l'irrégularité à gauche du résultat nul est peu prononcée. Mais elle apparaît tout de même comme l'une des plus importantes parmi les irrégularités négatives puisque le « pic » sur le graphe est l'un des plus « hauts ».

Tableau 160 : Distribution des irrégularités en 1992 ; ajustement Pareto2



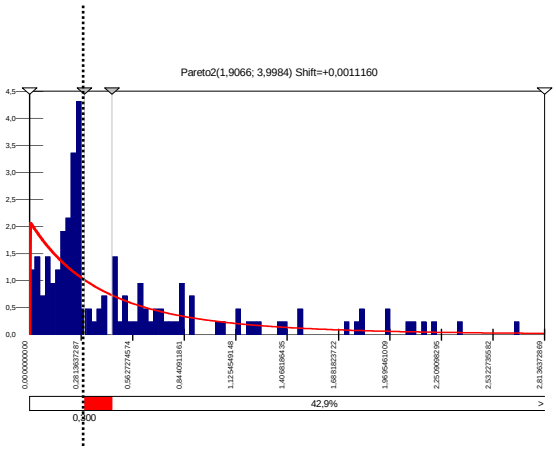
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,36114	0,26935
Max	0,45292	0,36114
Observé	14	32
Ajusté	11,19	13,85
Irrégularité	2,81	18,15
soit	**25,1%**	**131,0%**

En conclusion, l'étude de ce graphique est très intéressante. Dans les méthodes non paramétriques, les auteurs se contentaient de mesurer localement l'irrégularité, et estimaient par un test statistique sa significativité. La mesure paramétrique permet une représentation relative de l'irrégularité. Elle peut être comparée aux autres irrégularités qui se manifestent sur l'ensemble de la distribution. Autrement dit, ce n'est pas tant la significativité de l'irrégularité qu'il est intéressant de mesurer, que le fait que l'irrégularité au seuil du résultat nul est la plus prononcée de toutes les irrégularités observées sur la distribution.

Année 1993

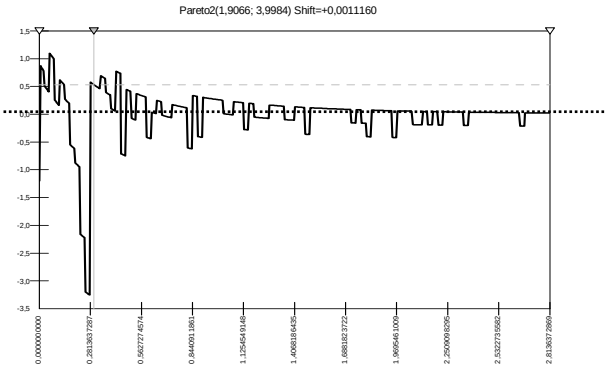
Tableau 161 : Ajustement de la demi-distribution gauche 1993 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(1,9066; 3,9984; RiskShift(0,0011160))
Shift	N/A	0,001115962
Parameter 1	N/A	1,906586567
Parameter 2	N/A	3,998408858
Minimum	0,001116	0,001116
Maximum	5,3851	+Infinity
Mean	0,63451	0,63698
Mode	0,27005	0,001116
Median	0,27571	0,36201
Std. Deviation	0,8304	0,89943
Variance	0,68491	0,80897
Skewness	2,9726	7,0787
Kurtosis	14,1017	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,1151
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	41,03
P Value	N/A	0,000048495
Rank	N/A	1



Position du résultat nul	0,281363729

Distribution des irrégularités



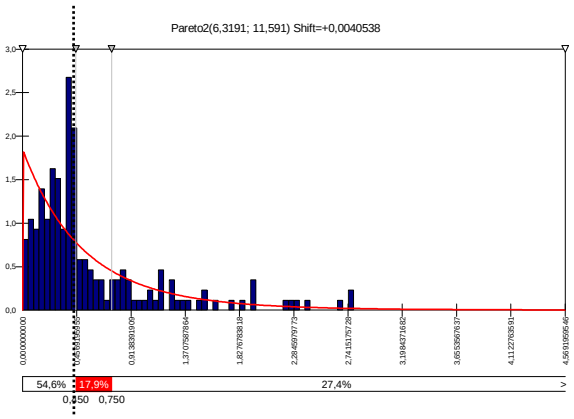
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,28136	0,2001976995
Max	0,36253	0,28136
Observé	5	38
Ajusté	11,6	13,96
Irrégularité	-6,6	
Soit	**-56,9%**	**172,2%**

L'irrégularité autour du résultat nul est observable graphiquement (seuil de 0,2814 sur la distribution ajustée), conformément à la théorie. L'étude de la distribution des irrégularités montre que en 1993, l'irrégularité à droite du résultat nul (172 %) est la plus forte de toute la distribution. L'irrégularité à gauche du résultat nul (-57 %) est moins prononcée, même si elle est parmi les deux plus importantes. Ces résultats son conformes à la théorie.

Année 1994

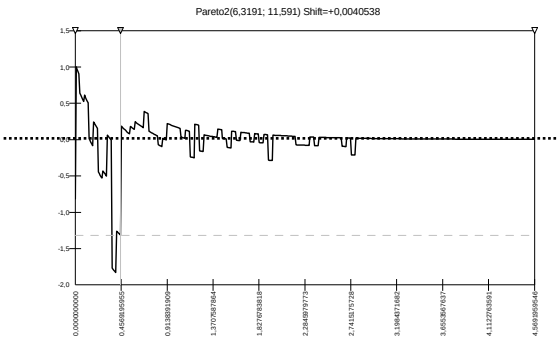
Tableau 162 : Ajustement de la demi-distribution gauche 1994 par une loi Pareto2

	Input	Pareto2
Function		=RiskPareto2(6,3191; 11,591; RiskShift(0,0040538))
Shift		0,004053845
Parameter 1		6,319122681
Parameter 2		11,5906479
Minimum	0,0040538	0,0040538
Maximum	6,1555	+Infinity
Mean	0,60201	0,60072
Mode	0,078758	0,0040538
Median	0,39328	0,39348
Std. Deviation	0,69225	0,65594
Variance	0,47667	0,43026
Skewness	3,7433	2,6664
Kurtosis	25,0594	15,8598
A-D Test Value		+Infinity
P Value		N/A
Rank		1
K-S Test Value		0,1126
P Value		N/A
Rank		1
Chi-Sq Test Value		54,39
P Value		1,1001E-06
Rank		1



Position du résultat nul	0,456919595

Distribution des irrégularités



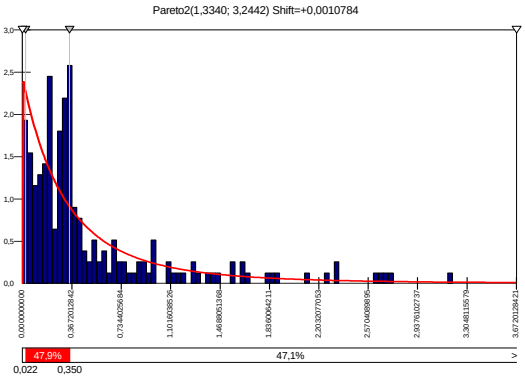
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,45692	0,36508
Max	0,54875	0,45692
Observé	10	41
Ajusté	12,18	14,45
Irrégularité	-2,18	26,55
Soit	**-17,9%**	**183,7%**

En 1994, l'irrégularité au dessus du résultat nul (184 %) est la plus importante de toute la distribution. L'irrégularité en dessous du résultat nul (-18 %) est très peu prononcée mais reste négative, conformément à la théorie.

Année 1995

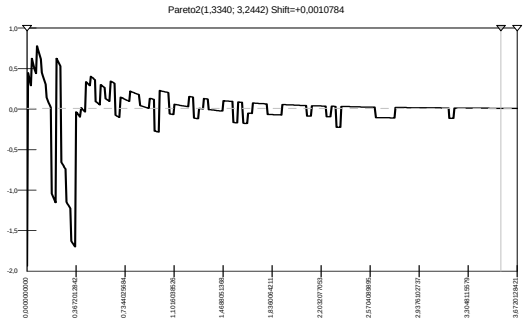
Tableau 163 : Ajustement de la demi-distribution gauche 1995 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(1,3340; 3,2442; RiskShift(0,0010784))
Shift	N/A	0,001078407
Parameter 1	N/A	1,333960423
Parameter 2	N/A	3,244197602
Minimum	0,0010784	0,0010784
Maximum	7,1837	+Infinity
Mean	0,59495	0,59548
Mode	0,026761	0,0010784
Median	0,31838	0,31883
Std. Deviation	0,87047	0,95982
Variance	0,75413	0,92126
Skewness	3,7836	21,5266
Kurtosis	22,4268	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,09213
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	46,84
P Value	N/A	0,000020394
Rank	N/A	1



Position du résultat nul	0,367201284

Distribution des irrégularités



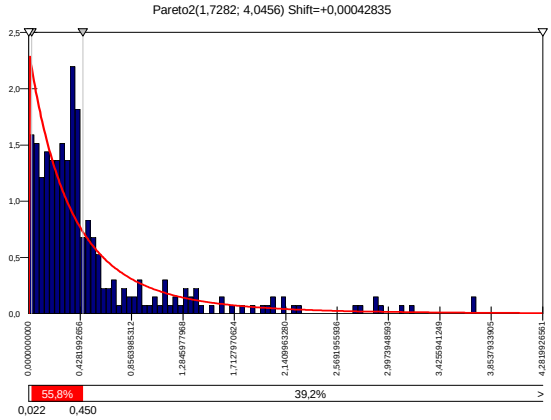
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,28633	0,3672
Max	0,3672	0,44807
Observé	41	13
Ajusté	16,46	13,44
Irrégularité	24,54	-0,44
Soit	**149,1%**	**-3,3%**

En 1995, l'irrégularité au dessus du résultat nul (149 %) est la plus importante de la distribution. L'irrégularité en dessous (-3 %) est quasi inexistante sur un intervalle de largeur 0,005. Il apparaît cependant qu'il y a une irrégularité négative en dessous du résultat nul, mais elle n'apparaît que si la largeur d'observation est agrandie. Ce constat est développé dans la section IX-B.

Année 1996

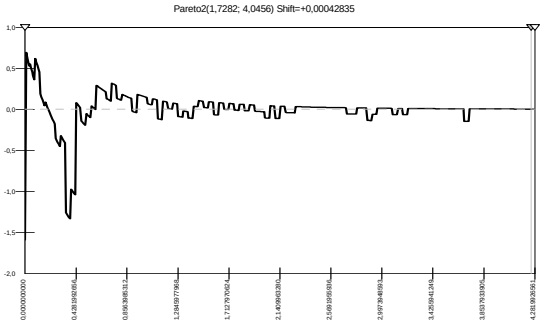
Tableau 164 : Ajustement de la demi-distribution gauche 1996 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(1,7282; 4,0456; RiskShift(0,00042835))
Shift	N/A	0,000428347
Parameter 1	N/A	1,728216785
Parameter 2	N/A	4,045554644
Minimum	0,00042835	0,00042835
Maximum	7,8298	+Infinity
Mean	0,57373	0,56788
Mode	0,44819	0,00042835
Median	0,34641	0,32341
Std. Deviation	0,84096	0,79802
Variance	0,70492	0,63684
Skewness	4,5534	6,8629
Kurtosis	32,2526	1756,2314
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,08339
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	52,58
P Value	N/A	0,000016625
Rank	N/A	1



Position du résultat nul	0,428199266

Distribution des irrégularités



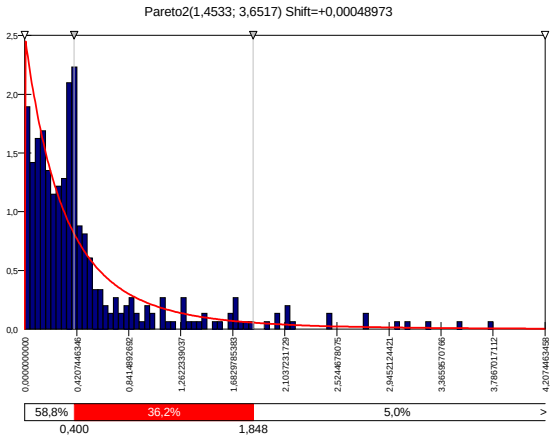
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,35304	0,4282
Max	0,4282	0,50336
Observé	43	16
Ajusté	19,43	16,3
Irrégularité	23,57	-0,3
Soit	**121,3%**	**-1,8%**

En 1996, l'irrégularité au dessus du résultat nul (121 %) apparaît comme la plus importante de la distribution. L'irrégularité en dessous du résultat nul (-2 %) est quasi inexistante si l'on se limite à une largeur d'observation de 0,005. Elle se manifeste en revanche si la largeur d'observation est élargie.

Année 1997

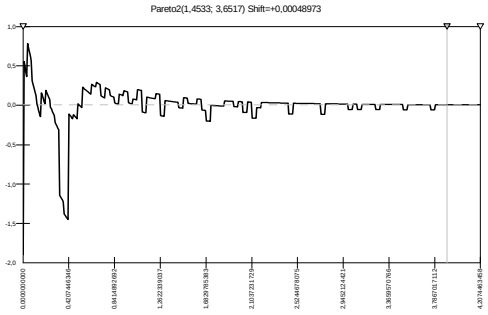
Tableau 165 : Ajustement de la demi-distribution gauche 1997 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(1,4533; 3,6517; RiskShift(0,00048973))
Shift	N/A	0,000489726
Parameter 1	N/A	1,453328963
Parameter 2	N/A	3,651657436
Minimum	0,00048973	0,00048973
Maximum	9,394	+Infinity
Mean	0,5575	0,54857
Mode	0,13620	0,00048973
Median	0,34315	0,30428
Std. Deviation	0,88571	0,81495
Variance	0,78226	0,66415
Skewness	5,4317	9,6014
Kurtosis	44,8546	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,08557
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	74,74
P Value	N/A	7,0124E-09
Rank	N/A	1



Position du résultat nul	0,420744635

Distribution des irrégularités



Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,35752	0,42074
Max	0,42074	0,48397
Observé	47	21
Ajusté	18,54	15,85
Irrégularité	28,46	5,15
Soit	**153,5%**	**32,5%**

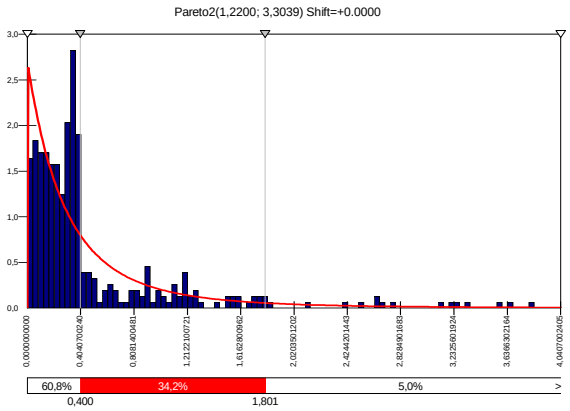
En 1997, l'irrégularité au dessus du résultat nul (153 %) apparaît graphiquement comme étant la plus importante de la distribution. L'irrégularité en dessous du résultat nul (32 %) est positive, ce qui n'est pas conforme à la théorie. C'est la seule année où un tel résultat se manifeste. La section IX-B montre qu'il est dépendant de la largeur d'intervalle retenue.

Année 1998

Tableau 166 : Ajustement de la demi-distribution gauche 1998 par une loi Pareto2

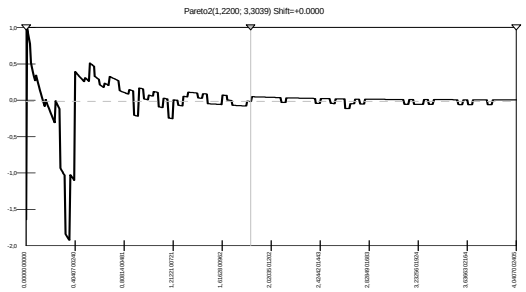
	Input	Pareto2
Function	N/A	
Shift	N/A	1E-10
Parameter 1	N/A	1,219968758
Parameter 2	N/A	3,303877382
Minimum	0.0000	0.0000
Maximum	7,1262	+Infinity
Mean	0,53265	0,52953
Mode	0,19338	0.0000
Median	0,30875	0,28478
Std. Deviation	0,80117	0,84291
Variance	0,64018	0,7105
Skewness	3,975	17,795
Kurtosis	24,0246	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,1193
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	113,7
P Value	N/A	0
Rank	N/A	1

=RiskPareto2(1,2200; 3,3039; RiskShift(0.0000))



Position du résultat nul	0,404070024

Distribution des irrégularités



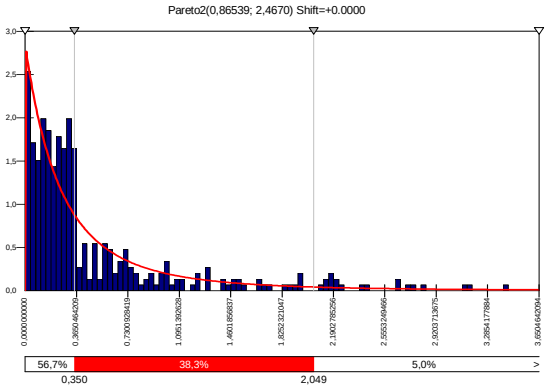
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,34628	0,40407
Max	0,40407	0,46186
Observé	46	9
Ajusté	18,63	15,98
Irrégularité	27,37	-6,98
Soit	**146,9%**	**-43,7%**

En 1998 , l'irrégularité au dessus du résultat nul (147 %) apparaît comme la plus importante (« trou » sur le graphique) de la distribution. L'irrégularité en dessous du résultat nul (-44 %) est, après l'irrégularité au sommet de la distribution, l'irrégularité négative (« pic » sur le graphique) la plus prononcée. Ces résultats sont conformes à la théorie.

Année 1999

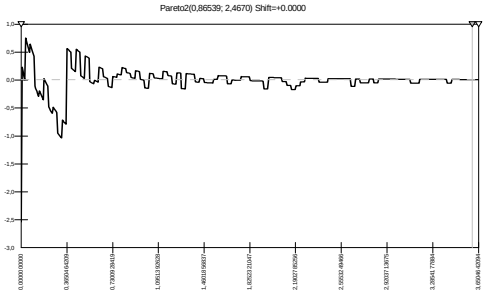
Tableau 167 : Ajustement de la demi-distribution gauche 1999 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(0,86539; 2,4670; RiskShift(0.0000))
Shift	N/A	1E-10
Parameter 1	N/A	0,865387943
Parameter 2	N/A	2,46696521
Minimum	0.0000	0.0000
Maximum	5,2778	+Infinity
Mean	0,57096	0,58992
Mode	0,20802	0.0000
Median	0,27072	0,28075
Std. Deviation	0,83754	1,3559
Variance	0,69971	1,8385
Skewness	2,926	N/A
Kurtosis	12,7387	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,08572
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	45,66
P Value	N/A	0,00055331
Rank	N/A	1



Position du résultat nul	0,365046421

Distribution des irrégularités



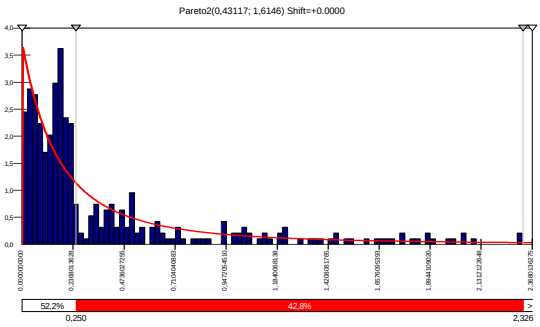
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,31226	0,36505
Max	0,36505	0,41783
Observé	35	8
Ajusté	19,13	16,48
Irrégularité	15,87	-8,48
Soit	**83,0%**	**-51,5%**

L'irrégularité au dessus du résultat nul (83 %) apparaît graphiquement comme la plus importante de la distribution, même si son amplitude est plus limitée que les années précédentes. L'irrégularité en dessous du résultat nul (-51 %) est en revanche plus importante que les années précédentes. Conformément à la théorie, c'est la plus importante des irrégularités négatives.

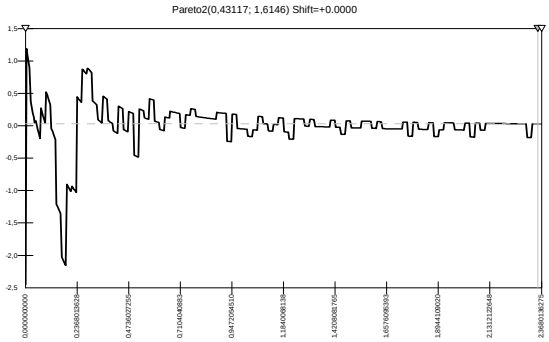
Année 2000

Tableau 168 : Ajustement de la demi-distribution gauche 2000 par une loi Pareto2

	Input	Pareto2		
Function	N/A		=RiskPareto2(0,43117; 1,6146; RiskShift(0.0000))	
Shift	N/A	1E-10		
Parameter 1	N/A	0,431166402		
Parameter 2	N/A	1,614618945		
Minimum	0.0000	0.0000		
Maximum	6,9917	+Infinity		
Mean	0,58764	0,70152		
Mode	0,027071	0.0000		
Median	0,19217	0,23118		
Std. Deviation	0,97673	N/A		
Variance	0,95159	N/A		
Skewness	3,1829	N/A		
Kurtosis	15,2255	N/A		
A-D Test Value	N/A	+Infinity		
P Value	N/A	N/A		
Rank	N/A	1		
K-S Test Value	N/A	0,09324		
P Value	N/A	N/A		
Rank	N/A	1		
Chi-Sq Test Value	N/A	87,64	Position du résultat nul	0,236801363
P Value	N/A	0		
Rank	N/A	1		



Distribution des irrégularités



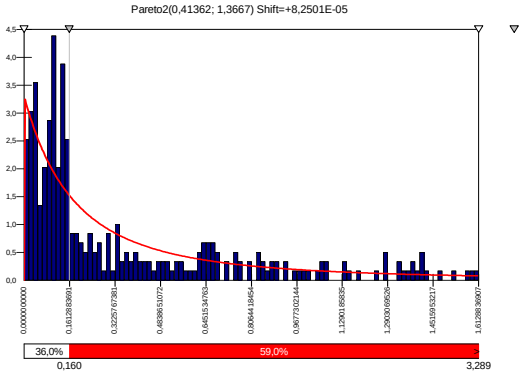
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,19462	0,2368
Max	0,2368	0,27899
Observé	35	9
Ajusté	21,7	18,39
Irrégularité	13,3	-9,39
Soit	**61,3%**	**-51,1%**

En 2000, la distribution semble illustrer parfaitement la théorie. La distribution est graphiquement bien ajustée, et les irrégularités autour du résultat nul sont les plus prononcées de toute la distribution.

Année 2001

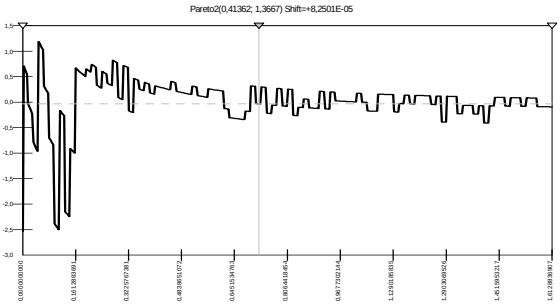
Tableau 169 : Ajustement de la demi-distribution gauche 2001 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(0,41362; 1,3667; RiskShift(8,2501E-05))
Shift	N/A	8,25005E-05
Parameter 1	N/A	0,413616898
Parameter 2	N/A	1,366701133
Minimum	6,1698	+Infinity
Maximum	0,71977	1,128
Mean	0,00014228	0,000082501
Mode	0,21934	0,27331
Median	1,0362	N/A
Std. Deviation	1,0708	N/A
Variance	2,2372	N/A
Skewness	8,3801	N/A
Kurtosis	N/A	+Infinity
A-D Test Value	N/A	N/A
P Value	N/A	1
Rank	N/A	0,09693
K-S Test Value	N/A	N/A
P Value	N/A	1
Rank	N/A	61,51
Chi-Sq Test Value	N/A	1,1641E-06
P Value	N/A	1
Rank		



Position du résultat nul	0,161288369

Distribution des irrégularités



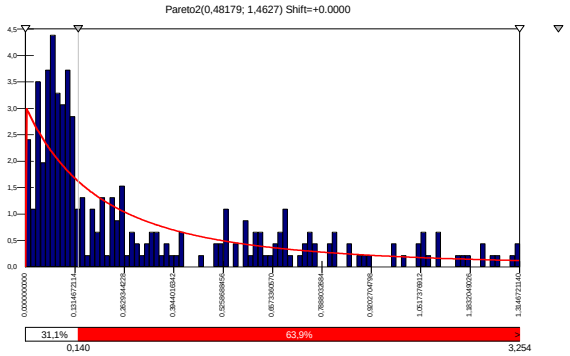
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,12366	0,16129
Max	0,16129	0,19892
Observé	42	13
Ajusté	22,69	19,43
Irrégularité	19,31	-6,43
Soit	**85,1%**	**-33,1%**

En 2001, l'irrégularité à droite du résultat nul est la plus forte de toute la distribution. L'irrégularité à gauche du résultat nul est moins prononcée en « hauteur », mais elle est observable sur une largeur plus importante. Ces résultats sont conformes à la théorie.

Année 2002

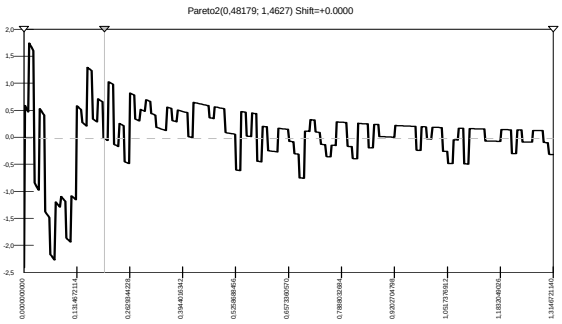
Tableau 170 : Ajustement de la demi-distribution gauche 2002 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(0,48179; 1,4627; RiskShift(0.0000))
Shift	N/A	1E-10
Parameter 1	N/A	0,481788385
Parameter 2	N/A	1,462679141
Minimum	0.0000	0.0000
Maximum	5,4989	+Infinity
Mean	0,73036	1,0413
Mode	0,026682	0.0000
Median	0,24755	0,29207
Std. Deviation	1,0251	N/A
Variance	1,0478	N/A
Skewness	2,239	N/A
Kurtosis	8,4039	N/A
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,1034
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	74,45
P Value	N/A	7,8697E-09
Rank	N/A	1



Position du résultat nul	0,131467211

Distribution des irrégularités



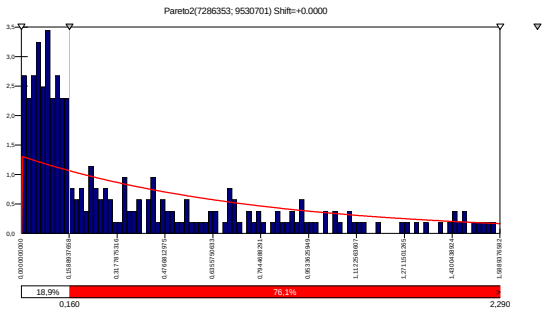
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,09316	0,13147
Max	0,13147	0,16977
Observé	42	11
Ajusté	24,12	20,68
Irrégularité	17,88	-9,68
soit	**74,1%**	**-46,8%**

En 2002, les irrégularité autour du résultat nul sont très prononcées, et conformes à la théorie.

Année 2003

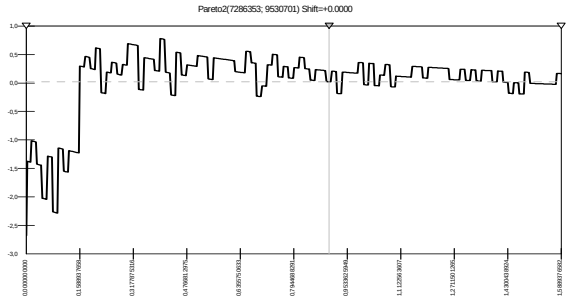
Tableau 171 : Ajustement de la demi-distribution gauche 2003 par une loi Pareto2

	Input	Pareto2
Function	N/A	=RiskPareto2(7286353; 9530701; RiskShift(0.0000))
Shift	N/A	1E-10
Parameter 1	N/A	7286352,897
Parameter 2	N/A	9530701,3
Minimum	0.0000	0.0000
Maximum	5,2765	+Infinity
Mean	0,76451	0,76451
Mode	0,11276	0.0000
Median	0,27142	0,52992
Std. Deviation	1,0363	0,76451
Variance	1,0706	0,58448
Skewness	1,9975	2
Kurtosis	6,8978	9
A-D Test Value	N/A	+Infinity
P Value	N/A	N/A
Rank	N/A	1
K-S Test Value	N/A	0,2355
P Value	N/A	N/A
Rank	N/A	1
Chi-Sq Test Value	N/A	140,3
P Value	N/A	0
Rank	N/A	1



Position du résultat nul	0,158893766

Distribution des irrégularités



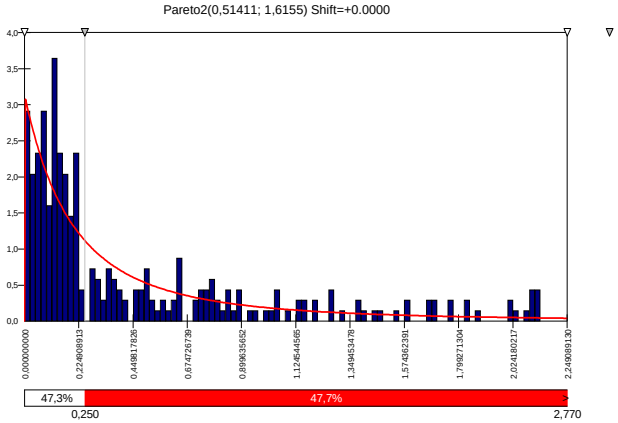
Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,11584	0,15889
Max	0,15889	0,20195
Observé	34	9
Ajusté	15,48	14,63
Irrégularité	18,52	-5,63
Soit	**119,6%**	**-38,5%**

En 2003, l'irrégularité est observable graphiquement. Cependant, il apparaît aussi que la distribution est globalement assez irrégulière. L'irrégularité autour du résultat nul se prolonge très longtemps, à droite comme à gauche. L'étude de la distribution des erreurs montre que les signes des irrégularités sont conformes à la théorie.

Année 2004

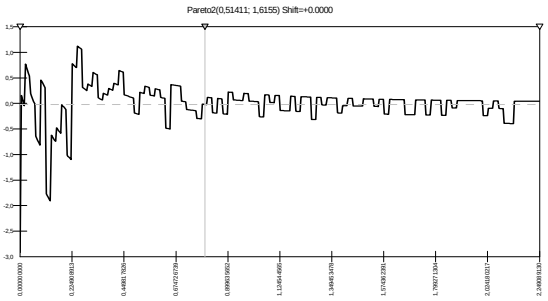
Tableau 172 : Ajustement de la demi-distribution gauche 2004 par une loi Pareto2

	Input	Pareto2
Function		
Shift		
Parameter 1		
Parameter 2		
Minimum		
Maximum		
Mean		
Mode		
Median		
Std. Deviation		
Variance		
Skewness		
Kurtosis		
A-D Test Value		
P Value		
Rank		
K-S Test Value		
P Value		
Rank		
Chi-Sq Test Value		
P Value		
Rank		



Position du résultat nul :	0,224908913

Distribution des irrégularités



Intervalle non transposé	[-0,005 ; 0]	[0 ; 0,005]
Min	0,22491	0,181
Max	0,26881	0,22491
Observé	3	26
Ajusté	15,11	17,65
Irrégularité	-12,11	8,35
Soit	**-80,1%**	**47,3%**

Les irrégularités autour du résultat nul sont les plus importantes de toute la distribution, conformément à la théorie.

Annexe 14 : Mesure de la largeur maximale année par année

Les mesures des largeurs maximales présentées ci-dessous détaillent les calculs effectués dans le cadre de la section IX-B (page 325). Les calculs sont présentés année par année.

Largeur max en 1992

L'irrégularité à gauche du seuil en 1992 est peu prononcée. Les mesures effectuées dans le tableau ci-dessous montrent que, même si elle est négative de manière très persistante, l'irrégularité ne dépasse jamais -17 %, et demeure donc très en deçà du seuil de significativité (-40 %) retenu. En 1992, une valeur nulle est donc attribuée à la largeur de l'irrégularité (et par conséquent la hauteur maximale) à gauche du seuil.

Tableau 173 : Irrégularité gauche (en 1992) mesurée en élargissant l'intervalle d'étude

Bornes de l'intervalle	Distribution du résultat	De à	-0,005 0	-0,01 0	-0,015 0	-0,02 0	-0,025 0	-0,03 0	-0,04 0	-0,05 0	-0,06 0	-0,07 0	-0,1 0
	Distribution ajustée	Soit de à	0,361135 0,452922	0,361135 0,544708	0,361135 0,636494	0,361135 0,728280	0,361135 0,820067	0,361135 0,911853	0,361135 1,095425	0,361135 1,278998	0,361135 1,462570	0,361135 1,646143	0,361135 2,19686
		Observé	14	19	27	34	35	39	43	47	50	51	53
		Ajusté	11,19	20,28	27,71	33,82	38,87	43,07	49,51	54,08	57,37	59,78	63,92
		écart	2,81	-1,28	-0,71	0,18	-3,87	-4,07	-6,51	-7,08	-7,37	-8,78	-10,92
		relatif	25%	-6%	-3%	1%	-10%	-9%	-13%	-13%	-13%	-15%	**-17%**

L'irrégularité à droite du seuil résultat nul (voir Tableau 174) est supérieure à 40 % jusqu'à une largeur maximale de 0,012. Le résultat s'élève alors à presque[201] 1,2 % de l'actif total. Sur cet intervalle, la distribution théorique prévoit 39,18 entreprises, alors que la distribution réelle en présente 56. Autrement dit, 16,8 entreprises sont en surnombre, soit 5,77 % du total des 291 entreprises présentes dans la population en 1992.

[201] Plus précisément, puisque nous étudions la distribution de la variable Résultat mise à l'échelle par (Actif Total – Résultat), si R/(A-R)=0,012 alors R=(0,012/1,012) x A soit R=0,01186 x A.

Tableau 174 : Irrégularité droite (en 1992) mesurée en élargissant l'intervalle d'étude

Bornes de l'intervalle	Distribution du résultat	De…	0	0	0	**0**	0	0	0	**0**	0	0	0
		À …	0,001	0,005	0,006	**0,007**	0,008	0,01	0,011	**0,012**	0,013	0,014	0,015
	Distribution ajustée	Soit de	0,342778	0,26934	0,250992	0,232635	0,214277	0,177563	0,159206	0,140848	0,122491	0,104134	0,085777
		à	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135	0,361135
		Observé	7	32	38	44	46	50	52	56	58	66	67
		Ajusté	2,54	13,86	17,01	20,3	23,74	31,11	35,05	39,18	43,5	48,03	52,78
		écart	4,46	18,14	20,99	**23,7**	22,26	18,89	16,95	**16,82**	14,5	17,97	14,22
		relatif	176%	131%	123%	117%	94%	61%	48%	**43%**	33%	37%	27%

La hauteur maximum est atteinte pour une largeur de 0,007 avec un surnombre de 23,7 entreprises, soit 23,7 / 291 = 8,14 % de l'échantillon. Elle constitue « l'irrégularité maximum » à droite du seuil.

Ce résultat appelle le commentaire suivant : le surnombre (8 % de l'échantillon) à droite du seuil n'est pas compensé par un sous-effectif à gauche du seuil. Autrement dit, la présomption de manipulation comptable de petite envergure, permettant à des entreprises qui ratent de peu le seuil, de publier malgré tout un résultat conforme aux attentes, n'est pas confirmée en 1992. Si les 23 à 24 entreprises, en surnombre au dessus du seuil du résultat nul, manipulent leur résultat pour dépasser le seuil, alors rien ne permet d'affirmer que le montant manipulé est faible.

Largeur max en 1993

La lecture de la partie haute du Tableau 175 révèle que l'irrégularité à gauche du seuil est supérieure à -40 % jusqu'à une largeur maximale de 0,012. Sur cet intervalle, 11,73 entreprises sont en surnombre, soit 3,96 % des 296 entreprises présentes dans la population en 1993. Cette irrégularité correspond également à la hauteur maximale.

Tableau 175 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1993

Irrégularité gauche													
De	-0,001	-0,005	-0,009	-0,01	-0,011	**-0,012**	-0,013	-0,015	-0,02	-0,03	-0,05	-0,07	-0,1
à	0	0	0	0	0	**0**	0	0	0	0	0	0	0
Soit de	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,281363	0,28136
à	0,297596	0,362529	0,427462	0,443695	0,459928	0,476162	0,492395	0,524861	0,606027	0,768359	1,09302	1,417688	1,9047
Observé	5	5	10	10	12	13	17	18	24	34	45	52	58
Ajusté	11,6	11,59	19,49	21,3	23,04	24,73	26,36	29,46	36,36	47,25	61,32	69,48	76,24
écart	-6,6	-6,59	-9,49	-11,3	-11,04	**-11,73**	-9,36	**-11,46**	-12,36	-13,25	-16,32	-17,48	-18,24
relatif	-57%	-57%	-49%	-53%	-48%	**-47%**	-36%	**-39%**	-34%	-28%	-27%	-25%	-24%

Irrégularité droite													
De	0	0	0	0	**0**	0	0	0	0	0	0	**0**	0
à	0,001	0,003	0,005	0,006	**0,007**	0,008	0,009	0,01	0,011	0,012	0,013	**0,014**	0,015
Soit de	0,2651305	0,23266411	0,2001977	0,18396449	0,16773128	0,15149808	0,13526487	0,11903167	0,10279846	0,08656525	0,07033205	0,0540988	0,03786564
à	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637	0,2813637
Observé	12	27	38	45	50	52	54	56	59	62	67	67	69
Ajusté	2,59	8,06	13,97	17,1	20,35	23,73	27,25	30,91	34,72	38,69	42,82	47,12	51,58
écart	9,41	18,94	24,03	27,9	**29,65**	28,27	26,75	25,09	24,28	23,31	24,18	**19,88**	17,42
relatif	363%	235%	172%	163%	**146%**	119%	98%	81%	70%	60%	56%	**42%**	34%

L'irrégularité à droite du seuil est supérieure à 40 % jusqu'à une largeur maximale de 0,014. Sur cet intervalle, 19,88 entreprises sont en surnombre, soit 6,72 % des 296 entreprises présentes dans la population en 1993. La hauteur maximale est atteinte pour une largeur de 0,007 avec 29,65 soit 10,02 % des 296 entreprises.

Il est intéressant de constater que le sous-effectif de près de 12 entreprises constaté à gauche sur une largeur de 0,07 ne suffit pas à compenser le surnombre de près de 30 entreprises observé à droite. Près de 18 entreprises, soit 6 % de l'échantillon total, en surnombre à droite du seuil, ne proviennent pas de la proximité gauche du seuil.

Largeur max en 1994

L'irrégularité à gauche du résultat nul est peu prononcée, et jamais supérieure au seuil retenu de 40 %. Elle est cependant de signe négatif. Il y a donc un sous-effectif non significatif.

Tableau 176 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1994

Irrégularité gauche														
De	-0,005	-0,01	-0,013	-0,014	-0,015	-0,016	-0,017	-0,02	-0,025	-0,03	-0,04	-0,05	-0,07	-0,1
à	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Soit de	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,4569
à	0,548754	0,640590	0,695691	0,714058	0,732425	0,750792	0,769159	0,824260	0,91609	1,007930	1,191601	1,375271	1,742612	2,2936
Observé	10	17	20	21	21	24	24	27	34	36	44	49	54	61
Ajusté	12,18	22,47	27,88	29,56	31,19	32,77	34,3	38,59	44,88	50,24	58,74	64,98	73,05	79,06
écart	-2,18	-5,47	-7,88	-8,56	-10,19	-8,77	-10,3	-11,59	-10,88	-14,24	-14,74	-15,98	-19,05	-18,06
relatif	-18%	-24%	-28%	-29%	-33%	-27%	-30%	-30%	-24%	-28%	-25%	-25%	-26%	

Irrégularité droite													
De	0	0	0	0	0	0	0	0	0	0	0	0	0
à	0,005	0,01	0,011	0,012	0,013	0,014	0,015	0,016	0,017	0,018	0,019	0,02	0,021
Soit de	0,365084	0,27324	0,254882	0,236515	0,218148	0,199780	0,181413	0,163046	0,144679	0,126312	0,107945	0,089578	0,07121
à	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919	0,456919
Observé	41	62	71	74	81	85	86	91	96	99	103	105	111
Ajusté	14,45	31,63	35,45	39,4	43,49	47,73	52,12	56,67	61,39	66,28	71,35	76,58	82,05
écart	26,55	30,37	35,55	34,6	**37,51**	37,27	33,88	34,33	34,61	32,72	**31,65**	28,42	28,95
relatif	184%	96%	100%	88%	**86%**	78%	65%	61%	56%	49%	**44%**	37%	35%

L'irrégularité à droite est supérieure à 40 % jusqu'à 0,019. Cela concerne alors 31,65 entreprises soit 8,42 % des 376 entreprises présentes en 1994. L'irrégularité maximale à droite du seuil est observable à une largeur de 0,013 avec 37,51 entreprises, soit 9,98 % de l'échantillon. Ce surnombre n'est pas compensé par les sous-effectifs à gauche du seuil.

Largeur max en 1995

L'irrégularité à gauche du résultat nul est peu prononcée, et jamais supérieure au seuil retenu de 40 %. Elle est cependant de signe négatif. Il y a un sous-effectif non significatif.

Tableau 177 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1995

Irrégularité gauche											
De	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,04	-0,05	-0,06	-0,07	-0,1
À	0	0	0	0	0	0	0	0	0	0	0
Soit de	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012
à	0,4480736	0,5289459	0,6098185	0,6906907	0,7715629	0,8524351	1,0141796	1,1759240	1,3376695	1,4994139	1,98465
Observé	13	18	25	31	36	39	47	51	55	58	66
Ajusté	13,44	24,53	33,75	41,47	47,99	53,53	62,33	68,89	73,88	77,74	85,09
écart	-0,44	-6,53	-8,75	-10,47	-11,99	-14,53	-15,33	-17,89	-18,88	-19,74	-19,09
relatif	-3%	-27%	-26%	-25%	-25%	-27%	-25%	-26%	-26%	-25%	

Irrégularité droite													
De	0	0	0	0	0	**0**	0	**0**	0	0	0	0	0
à	0,005	0,007	0,009	0,01	0,011	**0,013**	0,015	**0,016**	0,017	0,018	0,019	0,02	0,021
Soit de	0,2863289	0,2539800	0,2216311	0,2054566	0,1892821	0,1569332	0,1245843	0,1084098	0,0922353	0,0760609	0,0598864	0,0437119	0,0275375
à	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012	0,3672012
Observé	41	51	56	65	71	85	90	97	102	105	108	113	119
Ajusté	16,46	24,08	32,38	36,81	41,45	51,38	62,25	68,1	74,22	80,64	87,38	94,46	101,92
écart	24,54	26,92	23,62	28,19	29,55	**33,62**	27,75	28,9	27,78	24,36	20,62	18,54	17,08
relatif	149%	112%	73%	77%	71%	65%	45%	**42%**	37%	30%	24%	20%	17%

L'irrégularité à droite est significative jusqu'à une largeur de 0,016. à ce niveau, près de 29 entreprises sont en surnombre soit 6,85 % des 422 entreprises de l'échantillon en 1995. L'irrégularité maximale est atteinte avec une largeur de 0,013. Près de 34 entreprises, soit 7,97 % de la population, sont en surnombre à droite du seuil du résultat nul en 1995. Cette irrégularité n'est pas compensée par les sous-effectifs à gauche du seuil.

Largeur max en 1996

L'irrégularité à gauche du résultat nul est peu prononcée, et jamais supérieure au seuil retenu de 40 %. Elle est cependant de signe négatif.

Tableau 178 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1996

Irrégularité gauche											
De	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,04	-0,05	-0,06	-0,07	-0,1
à	0	0	0	0	0	0	0	0	0	0	0
Soit de	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,4282
à	0,5033617	0,5785241	0,6536865	0,7288490	0,8040114	0,8791738	1,0294987	1,1798236	1,3301485	1,4804733	1,93145
Observé	16	34	39	46	50	53	60	67	74	80	86
Ajusté	16,3	30,05	41,71	51,65	60,17	67,51	79,36	88,36	95,28	100,68	111,08
écart	-0,3	3,95	-2,71	-5,65	-10,17	-14,51	-19,36	-21,36	-21,28	-20,68	-25,08
relatif	-2%	13%	-6%	-11%	-17%	-21%	-24%	-24%	-22%	-21%	

Irrégularité droite													
De	0	0	0	0	0	0	0	**0**	0	**0**	0	0	0
À	0,005	0,009	0,01	0,014	0,015	0,016	0,017	**0,018**	0,019	**0,02**	0,021	0,022	0,025
Soit de	0,3530368	0,2929068	0,2778743	0,2177444	0,2027119	0,187679	0,172646	0,157614	0,142582	0,127549	0,1125170	0,0974845	0,0523870
À	0,4281992	0,4281992	0,4281992	0,4281992	0,4281992	0,428199	0,428199	0,428199	0,428199	0,428199	0,4281992	0,4281992	0,4281992
Observé	43	71	78	109	114	117	127	135	139	147	151	155	180
Ajusté	19,43	37,74	42,75	64,86	70,93	77,26	83,84	90,69	97,82	105,23	112,97	121,03	147,32
Écart	23,57	33,26	35,25	44,14	43,07	39,74	43,16	**44,31**	41,18	**41,77**	38,03	33,97	32,68
Relatif	121%	88%	82%	68%	61%	51%	51%	49%	42%	**40%**	34%	28%	22%

L'irrégularité à droite est significative jusqu'à une largeur de 0,02. À ce niveau, près de 42 entreprises sont en surnombre soit 6,78 % des 616 entreprises de l'échantillon en 1996. L'irrégularité maximum est atteinte avec une largeur de 0,018. Près de 44 entreprises, soit 7,19 % de la population, sont en surnombre à droite du seuil du résultat nul. Cette irrégularité n'est pas compensée par les sous-effectifs à gauche du seuil.

Largeur max en 1997

L'irrégularité à gauche du résultat nul est positive pour de très faibles intervalles (largeur inférieure à 0,01) mais devient négative lorsque l'on élargit l'intervalle d'observation. Elle reste cependant peu prononcée, et n'est jamais supérieure au seuil retenu de -40 %. Une valeur nulle est donc retenue pour la hauteur et la largeur maximales à gauche du seuil en 1997.

Tableau 179 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1997

Irrégularité gauche											
De	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,04	-0,05	-0,06	-0,07	-0.1
À	0	0	0	0	0	0	0	0	0	0	0
Soit de	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0,4207446	0.42074
À	0,4839660	0,5471875	0,6104089	0,6736304	0,7368518	0,8000733	0,9265162	1,0529591	1,1794020	1,3058448	1.68517
Observé	21	34	41	47	49	55	64	70	75	80	89
Ajusté	15,85	29,46	41,21	51,4	60,27	68,03	80,83	90,83	98,73	105,05	117.72
Écart	5,15	4,54	-0,21	-4,4	-11,27	-13,03	-16,83	-20,83	-23,73	-25,05	-28.72
Relatif	**32%**	**15%**	-1%	-9%	-19%	-19%	-21%	-23%	-24%	-24%	-24%

Irrégularité droite														
De	0	0	0	0	**0**	0	0	0	0	**0**	0	0	0	0
À	0,005	0,009	0,01	0,011	**0,013**	0,015	0,019	0,02	0,021	**0,022**	0,023	0,024	0,025	0,03
Soit de	0,35752318	0,30694602	0,29430174	0,28165745	0,2563688	0,23108029	0,18050313	0,16785884	0,15521455	0,1425702	0,12992597	0,11728168	0,10463739	0,04141595
À	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744	0,420744
Observé	47	76	83	89	101	109	134	138	145	156	163	168	174	209
Ajusté	18,53	35,69	40,35	45,17	55,31	66,17	90,32	96,92	103,76	110,85	118,21	125,85	133,77	178,31
Écart	28,47	40,31	42,65	43,83	**45,69**	42,83	43,68	41,08	41,24	**45,15**	44,79	42,15	40,23	30,69
Relatif	154%	113%	106%	97%	83%	65%	48%	42%	40%	**41%**	38%	33%	30%	17%

L'irrégularité à droite est très prononcée. Elle est significative jusqu'à une largeur de 0,022. À ce niveau, près de 45 entreprises sont en surnombre soit 6,43 % des 702 entreprises de l'échantillon en 1997. La hauteur maximale est de 45,69 entreprises observée avec une largeur de 0,013. Autrement dit, 6,5 % des entreprises manipulent leurs résultats pour éviter le seuil du résultat nul en 1997, et la manipulation marginale ne suffit pas à expliquer cette irrégularité.

Largeur max en 1998

L'irrégularité à gauche du résultat nul est nettement prononcée, et très persistante. Elle est significative jusqu'à une largeur de 0,073. L'irrégularité maximale à gauche du seuil en 1998 s'élève à 49,27 entreprises (pour une largeur de 0,06) soit 6,54 % des 753 entreprises de l'échantillon.

Tableau 180 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1998

Irrégularité gauche														
De	-0,005	-0,01	-0,03	-0,05	-0,059	**-0,06**	-0,061	-0,07	-0,072	**-0,073**	-0,074	-0,075	-0,08	-0,1
À	0	0	0	0	0	**0**	0	0	0	**0**	0	0	0	0
Soit de	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,404070	0,404070
À	0,461864	0,519659	0,750838	0,982017	1,086047	1,097606	1,109165	1,213195	1,236313	1,247872	1,259431	1,270990	1,328785	1,55996
Observé	9	16	29	48	52	52	54	64	66	66	67	68	71	75
Ajusté	15,98	29,77	69,21	92,93	100,49	101,27	101,98	107,99	109,14	109,71	110,27	110,82	112,92	121,69
Écart	-6,98	-13,77	-40,21	-44,93	-48,49	**-49,27**	-47,98	-43,99	-43,14	**-43,71**	-43,27	-42,82	-41,92	-46,69
Relatif	-44%	-46%	-58%	-48%	-48%	-49%	-47%	-41%	-40%	**-40%**	-39%	-39%	-37%	-38%

Irrégularité droite														
De	0	0	0	0	0	**0**	0	0	0	0	**0**	0	0	0
à	0,005	0,01	0,015	0,018	0,019	**0,02**	0,021	0,022	0,025	0,027	**0,028**	0,029	0,03	0,032
Soit de	0,346275	0,288480	0,230685	0,196009	0,184450	0,172891	0,16133	0,149773	0,115096	0,091978	0,080419	0,068860	0,05730	0,034183
à	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407	0,40407
Observé	46	96	132	151	158	167	170	175	201	215	222	233	241	252
Ajusté	18,63	40,47	66,42	84,13	90,46	97,01	103,8	110,84	133,57	150,2	159	168,14	177,64	197,79
écart	27,37	55,53	65,58	66,87	67,54	**69,99**	66,2	64,16	67,43	64,8	63	64,86	63,36	54,21
relatif	147%	137%	99%	79%	75%	72%	64%	58%	50%	43%	**40%**	39%	36%	27%

L'irrégularité à droite est très prononcée. Elle est significative jusqu'à une largeur de 0,028. À ce niveau, 63 entreprises sont en surnombre. L'irrégularité maximale est constatée sur une largeur de 0,02 avec 70 entreprises en surnombre, soit 9,3 % des 753 entreprises de l'échantillon en 1998. Au total, 21 entreprises en surnombre à droite du seuil ne proviennent pas de la proximité gauche du seuil.

Largeur max en 1999

L'irrégularité à gauche du résultat nul est nettement prononcée. Elle est supérieure à -40 % jusqu'à une largeur de 0,022, mais continue à être importante bien au-delà. L'irrégularité maximale à gauche du seuil s'élève à 24,15 entreprises (pour une largeur de 0,02) soit 3,03 % des 797 entreprises de l'échantillon.

Tableau 181 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1999

Irrégularité gauche																	
De	-0,001	-0,005	-0,01	-0,015	-0,018	-0,019	**-0,02**	-0,021	**-0,022**	-0,025	-0,026	-0,03	-0,04	-0,05	-0,06	-0,07	-0,1
À	0	0	0	0	0	0	**0**	0	**0**	0	0	0	0	0	0	0	0
Soit de	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504	0,36504
À	0,37560	0,41783	0,47061	0,52339	0,55506	0,56562	0,57618	0,58674	0,59729	0,62896	0,63952	0,68175	0,78732	0,89288	0,99845	1,10402	1,42073
Observé	0	8	14	22	26	28	30	33	35	40	40	45	61	66	76	79	90
Ajusté	3,49	16,48	30,77	43,23	49,95	52,08	54,15	56,17	58,14	63,76	65,55	72,27	86,58	98,04	107,34	114,98	131,13
écart	-3,49	-8,48	-16,77	-21,23	-23,95	-24,08	**-24,15**	-23,17	**-23,14**	-23,76	-25,55	-27,27	-25,58	-32,04	-31,34	-35,98	-41,13
relatif	-100%	-51%	-55%	-49%	-48%	-46%	-45%	-41%	**-40%**	-37%	-39%	-38%	-30%	-33%	-29%	-31%	

Irrégularité droite												
De	0	0	0	0	0	0	0	0	**0**	**0**	0	0
À	0,001	0,002	0,005	0,01	0,015	0,02	0,021	0,022	**0,023**	**0,024**	0,025	0,03
Soit de	0,354489	0,343932	0,312262	0,259477	0,206693	0,153909	0,143352	0,132795	0,122238	0,111682	0,101125	0,048340
à	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046	0,365046
Observé	4	13	35	74	111	148	153	164	175	179	181	220
Ajusté	3,6	7,31	19,13	41,47	67,77	98,97	105,9	113,09	120,55	128,29	136,33	181,47
écart	0,4	5,69	15,87	32,53	43,23	49,03	47,1	50,91	**54,45**	**50,71**	44,67	38,53
relatif	11%	78%	83%	78%	64%	50%	44%	45%	45%	**40%**	33%	21%

L'irrégularité à droite est supérieure à 40 % jusqu'à une largeur de 0,024. L'irrégularité maximale est atteinte à la largeur 0,023 avec 54,45 entreprises en surnombre, soit 6,83 % des 797 entreprises de l'échantillon en 1999. Au total, il y a une différence de 20 entreprises entre les deux irrégularités. Il faut élargir jusqu'à plus de 0,07 du coté gauche pour parvenir à diminuer de moitié seulement cette différence. Autrement dit, les montants manipulés pour éviter le seuil du résultat nul ne sont pas nécessairement de petite ampleur.

Largeur max en 2000

L'irrégularité à gauche du résultat nul est négative et supérieure au seuil retenu de -40 % jusqu'à une largeur de 0,026. Au-delà de cette largeur, l'irrégularité reste proche du seuil de 40 %. Cette valeur est cependant retenue comme largeur maximale, qui coïncide avec la largeur permettant d'observer la hauteur maximale : près de 30 entreprises, soit 29,85/791= 3,77 % de la population de 2000.

Tableau 182 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2000

Irrégularité gauche																
De	-0,005	-0,01	-0,015	-0,018	-0,02	-0,022	-0,024	-0,025	**-0,026**	-0,027	-0,029	-0,03	-0,04	-0,05	-0,06	-0,1
À	0	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0
Soit de	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680
À	0,27899	0,32117	0,36336	0,38867	0,40554	0,42241	0,43929	0,44773	0,45616	0,46460	0,48147	0,48991	0,57429	0,65866	0,74303	1,08052
Observé	9	15	24	26	32	38	40	41	42	45	48	48	64	73	78	96
Ajusté	18,39	34,13	47,72	55,02	59,55	63,87	67,96	69,92	71,85	73,72	77,33	79,06	94,4	106,71	116,76	143,07
Écart	-9,39	-19,13	-23,72	-29,02	-27,55	-25,87	-27,96	-28,92	**-29,85**	-28,72	-29,33	-31,06	-30,4	-33,71	-38,76	-47,07
Relatif	-51%	-56%	-50%	-53%	-46%	-41%	-41%	-41%	**-42%**	-39%	-38%	-39%	-32%	-32%	-33%	-33%

Irrégularité droite																
De	0	0	0	0	0	0	0	**0**	0	0	0	0	**0**	0	0	0
à	0,001	0,002	0,004	0,005	0,007	0,01	0,012	**0,013**	0,014	0,015	0,016	0,018	**0,019**	0,02	0,023	0,025
Soit de	0,22836	0,21992	0,20305	0,19461	0,17774	0,15242	0,13555	0,12711	0,11868	0,11024	0,10180	0,08493	0,07649	0,06805	0,04274	0,02587
à	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680	0,23680
Observé	5	12	26	35	67	94	111	119	123	131	137	152	156	161	196	210
Ajusté	4,05	8,24	17,05	21,69	31,48	47,57	59,36	65,6	72,1	78,86	85,9	100,9	108,89	117,23	144,62	165,13
écart	0,95	3,76	8,95	13,31	35,52	46,43	51,64	**53,4**	50,9	52,14	51,1	51,1	**47,11**	43,77	51,38	44,87
relatif	23%	46%	52%	61%	113%	98%	87%	81%	71%	66%	59%	51%	**43%**	37%	36%	27%

L'irrégularité à droite est très prononcée. Elle est significative jusqu'à une largeur de 0,019. À ce niveau, plus de 47 entreprises sont en surnombre soit 5,96 % des 791 entreprises de l'échantillon en 2000. La hauteur maximale de l'irrégularité se situe à une largeur de 0,013 et représente 53,4 entreprises, soit 6,75 % des entreprises de l'échantillon. Comparer les deux irrégularités conduit à penser qu'une petite manipulation du résultat pour éviter le seuil ne suffit pas à expliquer près de la moitié des cas irréguliers.

Par ailleurs, la notion de petite manipulation est à relativiser. La distance totale entre largeur maximale où l'irrégularité est sensible à gauche et à droite est de 0,019 + 0,026, soit 0,045. Si l'on retient une distance moyenne de 0,045/2 = 0,0225, alors ce que l'on appelle « petite manipulation » correspond à une manipulation de l'ordre de 2,25 % du total actif. La notion de « petite manipulation » est donc tout à fait relative.

Largeur max en 2001

L'irrégularité à gauche du résultat nul est peu significative, mais très persistante. Elle est supérieure au seuil retenu de -40 % jusqu'à une largeur de 0,061. Au-delà de cette largeur, l'irrégularité reste proche du seuil de 40 %. Cette valeur est cependant retenue comme largeur maximale, qui coïncide avec la largeur permettant d'observer la hauteur maximale : près de 55 entreprises, soit 55,12 / 734 = 7,51 % des 734 entreprises de la population de 2001.

Tableau 183 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2001

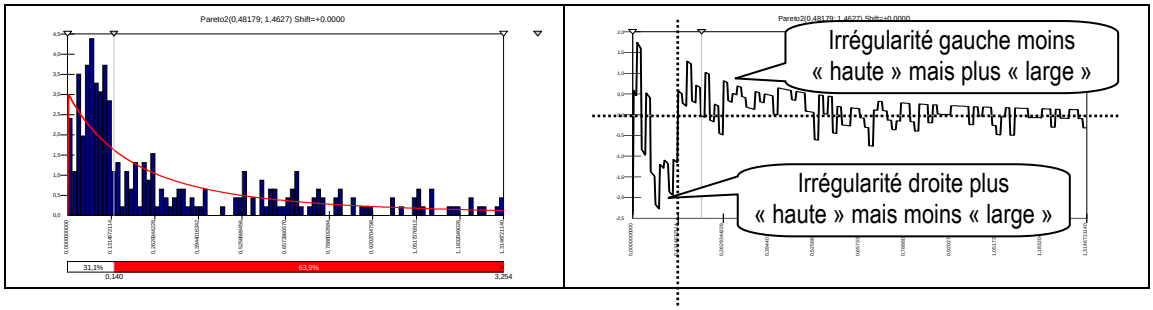
Irrégularité gauche															
De	-0,001	-0,009	-0,01	-0,011	-0,03	-0,035	-0,04	-0,05	-0,059	-0,06	**-0,061**	-0,065	-0,07	-0,08	-0,1
à	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0	0
Soit de	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128
à	0,168814943	0,229027535	0,236554109	0,244080683	0,387085589	0,	0,462351329	0,537617068	0,605356234	0,612882808	0,620409382	0,650516	0,688148548	0,763414288	0,91395
Observé	2	18	22	22	49		59	67	73	73	74	81	91	97	110
Ajusté	4,12	33,05	36,22	39,31	85,24		102,43	116,45	127	128,07	129,12	133,16	137,84	146,14	159,49
écart	-2,12	-15,05	-14,22	-17,31	-36,24	-40,29	-43,43	-49,45	-54	-55,07	**-55,12**	-52,16	-46,84	-49,14	-49,49
relatif	-51%	-46%	-39%	-44%	-43%		-42%	-42%	-43%	-43%	**-43%**	-39%	-34%	-34%	%

Irrégularité droite															
De	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
à	0,001	0,002	0,005	0,006	0,008	0,01	0,011	0,012	0,013	0,014	0,015	0,016	0,017	0,018	0,02
Soit de	0,153761	0,146235221	0,123655499	0,116128925	0,101075777	0,086022629	0,078496055	0,070969481	0,063442907	0,055916333	0,048389759	0,040863185	0,033336611	0,025810037	0,010756889
à	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128	0,16128
Observé	5	13	42	48	70	87	94	101	105	109	113	124	131	137	156
Ajusté	4,25	8,64	22,69	27,69	38,21	49,49	55,43	61,59	67,98	74,62	81,51	88,68	96,13	103,89	120,39
écart	0,75	4,36	19,31	20,31	31,79	37,51	38,57	**39,41**	37,02	34,38	31,49	35,32	34,87	33,11	35,61
relatif	18%	50%	85%	73%	83%	76%	70%	**64%**	54%	46%	39%	**40%**	36%	32%	30%

L'irrégularité à droite est modérée. Elle est significative jusqu'à une largeur de 0,014. À ce niveau, 34,38 entreprises sont en surnombre soit 4,68 % des 734 entreprises de l'échantillon de 2001. La hauteur maximale de l'irrégularité droite se situe à une largeur de 0,012 avec 39,41 entreprises excédentaires, soit 5,37 % des 734 entreprises composant la population de 2001. L'année 2001 constitue, avec 2004, l'une des deux seules années où l'irrégularité à gauche est, en valeur absolue, supérieure à l'irrégularité à droite.

Largeur max en 2002

Figure 136 : Représentation graphique de la demi-distribution gauche des résultats (transposés) en 2002



En 2002, l'irrégularité à gauche du résultat nul est moins prononcée que l'irrégularité à droite, mais elle persiste sur une largeur d'intervalle plus importante. Une particularité liée à 2002 vient du fait que l'irrégularité est supérieure au seuil retenu de -40 % jusqu'à une largeur de 0,009 seulement. Au-delà de cette largeur, l'irrégularité oscille entre -30 et -40 %, puis redevient supérieure à -40 % entre 0,045 et 0,059. Autrement dit, le seuil (« sévère ») de significativité fixé tend à sous-estimer largement l'irrégularité maximale à gauche du seuil en 2002.

Tableau 184 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2002

Irrégularité gauche																	
De	-0,001	-0,002	-0,005	-0,008	**-0,009**	-0,01	-0,02	-0,03	-0,04	-0,041	-0,045	-0,05	-0,055	**-0,059**	-0,06	-0,07	-0,1
à	0	0	0	0	**0**	0	0	0	0	0	0	0	0	**0**	0	0	0
Soit de	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146	0,13146
à	0,13912	0,14679	0,16977	0,19275	0,20042	0,20808	0,28469	0,36131	0,43792		0,47623	0,51454	0,55284	0,58349	0,59115	0,66777	0,89762
Observé	4	5	11	19	20	24	48	59	67		68	70	77	80	83	94	120
Ajusté	4,39	8,65	20,68	31,71	35,19	38,57	67,87	90,75	109,04		116,84	123,93	130,36	135,11	136,24	146,56	169,32
écart	-0,39	-3,65	-9,68	-12,71	**-15,19**	-14,57	-19,87	-31,75	-42,04	-43,67	-48,84	-53,93	-53,36	**-55,11**	-53,24	-52,56	-49,32
relatif	-9%	-42%	-47%	-40%	**-43%**	-38%	-29%	-35%	-39%		-42%	-44%	-41%	**-41%**	-39%	-36%	%

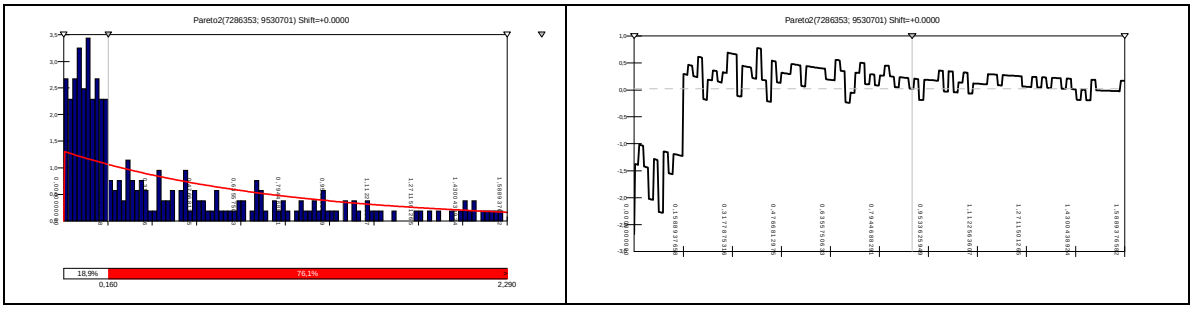
Irrégularité droite												
De	0	0	0	0	0	**0**	0	0	0	0	**0**	0
A	0,001	0,002	0,003	0,005	0,009	**0,01**	0,011	0,012	0,013	0,014	**0,015**	0,016
Soit de	0,1238057	0,116144025	0,108482017	0,093159 8	0,0625139	0,0548524	0,0471909	0,03952904	0,0318679	0,0242065	0,0165450	0,0088835
A	0,131467	0,1314670	0,1314670	0,131467	0,131467	0,131467	0,131467	0,1314670	0,131467	0,131467	0,131467	0,131467
Observé	6	16	26	42	84	96	99	105	107	122	126	129
Ajusté	4,52	9,19	14,01	24,12	46,47	52,55	58,85	65,38	72,15	79,17	86,46	94,03
Écart	1,48	6,81	11,99	17,88	37,53	**43,45**	40,15	39,62	34,85	42,83	39,54	34,97
relatif	33%	74%	86%	74%	81%	**83%**	68%	61%	48%	54%	**46%**	37%

La largeur de 0,009 est cependant retenue comme largeur maximale. Elle coïncide avec la largeur permettant d'observer la hauteur maximale : 15,19 entreprises, soit 2,19 % des 693 entreprises de la population de 2002. L'irrégularité à droite est beaucoup plus

prononcée. Elle apparaît significative jusqu'à une largeur de 0,015 représentant alors 39,54 entreprises. Elle culmine à la largeur de 0,01 avec 43,45 entreprises en surnombre, soit 5,71 % d'une population de 693. Une fois de plus, le sur nombre d'entreprises à droite n'est pas compensé par un sous-effectif à gauche. Pour annuler la somme des deux irrégularités, il faut élargir l'intervalle de gauche à 0,04 environ (42 entreprises manquantes). Le montant moyen manipulé par les entreprises faiblement déficitaires pour éviter le seuil est alors de (0,04 + 0,01) x ½ soit de l'ordre de 2,5 % du total actif.

Largeur max en 2003

Figure 137 : Représentation graphique de la demi-distribution gauche des résultats (transposés) en 2003



En 2003, plus encore qu'en 2002, la forme de la distribution est atypique (voir Figure 137), et l'irrégularité à gauche du résultat nul est peu significative, mais très persistante. La largeur maximale est fortement dépendante du niveau minimum fixé. En respectant le critère de -40 %, alors l'irrégularité n'est pas significative. Pourtant, elle apparaît très proche du seuil, atteignant plusieurs fois le niveau limite de -39 %. Il est remarquable qu'en élargissant l'intervalle au-delà de 0,06, l'irrégularité continue d'être sensible, et dépasse même les -40 %.

Tableau 185 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2003

	Irrégularité gauche												
De	-0,001	-0,002	-0,005	-0,008	-0,01	-0,015	-0,02	-0,03	-0,04	-0,05	-0,06	**-0,07**	-0,1
à	0	0	0	0	0	0	0	0	0	0	0	**0**	0
Soit de	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937	0,1588937
à	0,1675042	0,1761148	0,2019464	0,2277779	0,2449990	0,2880516	0,3311043	0,4172096	0,5033149	0,5894202	0,6755254	0,7616307	1,01995
Observé	3	4	9	14	20	31	34	47	61	70	76	87	107
Ajusté	2,99	5,95	14,63	23,03	28,46	41,54	53,9	76,62	96,93	115,08	131,29	145,78	180,6
Écart	0,01	-1,95	-5,63	-9,03	-8,46	-10,54	-19,9	-29,62	-35,93	-45,08	-55,29	**-58,78**	-73,6
Relatif	0%	-33%	-38%	-39%	-30%	-25%	-37%	-39%	-37%	-39%	-42%	**-40%**	-41%

Irrégularité droite									
De	0	0	0	0	0	0	0	0	**0**
À	0,001	0,002	0,005	0,01	0,015	0,016	0,017	0,018	**0,01845**
Soit de	0,150283237	0,141672708	0,115841122	0,072788477	0,029735833	0,021125304	0,012514776	0,003904247	0
À	0,158893766	0,158893766	0,158893766	0,158893766	0,158893766	0,158893766	0,158893766	0,158893766	0,158893766
Observé	6	14	34	77	112	120	127	134	138
Ajusté	3,03	6,09	15,48	31,86	49,19	52,77	56,39	60,06	61,74
écart	2,97	7,91	18,52	45,14	62,81	67,23	70,61	73,94	**76,26**
relatif	98%	130%	120%	142%	128%	127%	125%	123%	**124%**

L'année 2003 présente un problème technique spécifique pour mesurer l'irrégularité à droite du seuil. En effet, l'irrégularité étant située assez proche du sommet de la distribution, en transposant la demi-distribution des valeurs inférieures à la médiane, il devient impossible de sélectionner une largeur d'intervalle supérieure à la distance séparant le seuil et la médiane (c'est-à-dire le point zéro de la distribution transposée). Ici, la distance maximale est légèrement supérieure à 0,018[202]. Or en 2003, l'irrégularité à droite étant très prononcée et très persistante, elle est encore très supérieure à 40 % sur l'intervalle maximal. Il faut donc prendre la valeur maximale mesurable, sachant que la largeur maximale comme la hauteur maximale de l'irrégularité sont sans doute sous-estimées. L'irrégularité maximale représente donc au moins 76,26 entreprises soit 11,6 % d'une population totale de 657.

Le surnombre à droite du seuil n'est, une fois de plus, pas compensé par le sous-effectif à gauche du seuil. Il est intéressant de constater que ce constat est indépendant du seuil retenu pour reconnaître une irrégularité. En effet, à gauche du seuil, si un niveau minimum de 30 % est retenu, l'irrégularité maximale observée est de 73 entreprises (pour une largeur de 0,10), et il persiste un déficit d'entreprises manquantes à gauche du seuil pour compenser le sur nombre à droite. Autrement dit, en 2003, les entreprises qui évitent le résultat nul (soit près de 12 % de la population) ont manipulé en moyenne plus de (0,10 + 0,018) x ½ soit plus de 5,9 % de la valeur du total de leur actif. La présomption de manipulation marginale pour éviter un seuil est difficilement défendable.

[202] Largeur de 0,158 sur la distribution transposée et ajustée, donc 0,018 sur la distribution initiale.

Largeur max en 2004

L'irrégularité à gauche du résultat nul est négative et supérieure au seuil retenu de -40 % jusqu'à une largeur de 0,061. Même au-delà de cette largeur, l'irrégularité reste proche du seuil de 40 %. L'irrégularité maximale est observable à 0,06 pour 40,57 entreprises, soit 6,66 % d'une population de 609 entreprises en 2004.

Tableau 186 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2004

Irrégularité gauche															
De	-0,001	-0,005	-0,01	-0,02	-0,03	-0,04	-0,05	-0,059	**-0,06**	**-0,061**	-0,062	-0,063	-0,065	-0,067	-0,1
À	0	0	0	0	0	0	0	0	**0**	**0**	0	0	0	0	0
Soit de	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908	0,224908
À	0,233689	0,268814	0,312719	0,400529	0,48833	0,576149	0,663960	0,742989	0,751770	0,760551	0,769332	0,778113	0,795675	0,813237	1,10301
Observé	0	3	12	26	34	43	53	58	58	60	63	63	67	67	83
Ajusté	3,21	15,11	28,15	49,45	66	79,16	89,81	97,77	98,57	99,36	100,14	100,9	102,39	103,82	121,81
Écart	-3,21	-12,11	-16,15	-23,45	-32	-36,16	-36,81	-39,77	**-40,57**	-39,36	-37,14	-37,9	-35,39	-36,82	-38,81
Relatif	-100%	-80%	-57%	-47%	-48%	-46%	-41%	-41%	**-41%**	**-40%**	-37%	-38%	-35%	-35%	

Irrégularité droite											
De	0	0	0	0	0	0	**0**	0	0	0	0
À	0,001	0,002	0,005	0,01	0,015	0,016	**0,017**	0,018	0,019	0,02	0,025
Soit de	0,2161278	0,20734686	0,18100379	0,13709868	0,0931935	0,08441254	0,07563152	0,06685050	0,05806947	0,04928845	0,00538334
À	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089	0,2249089
Observé	3	10	26	55	90	98	106	112	118	123	159
Ajusté	3,31	6,72	17,65	38,48	63,33	68,88	74,65	80,64	86,88	93,37	130,21
Écart	-0,31	3,28	8,35	16,52	26,67	29,12	**31,35**	31,36	31,12	29,63	28,79
Relatif	-9%	49%	47%	43%	42%	42%	**42%**	39%	36%	32%	22%

En 2004, l'irrégularité à gauche est nettement plus prononcée que l'irrégularité à droite, ce qui est une exception dans la série chronologique. L'irrégularité droite demeure supérieure à 40 % jusqu'à une largeur de 0,017. L'irrégularité maximale s'élève à 31,35 entreprises, soit 5,15 % d'une population de 609. L'année 2004 est, avec 2001, l'une des deux seules années dans l'étude où les entreprises manquantes (à gauche du seuil) sont plus nombreuses que les entreprises surnuméraires (à droite du seuil).

Annexe 15 : Calcul du montant moyen annuel manipulé par les entreprises qui évitent une perte ; calcul par moyenne pondérée

Le tableau ci-dessous détaille les calculs présentés dans le paragraphe IX-B-2-d.2. détermination du montant moyen manipulé par moyenne pondérée de la page 337. Il permet de déterminer, par un calcul de moyenne pondérée, le montant annuel moyen manipulé par les entreprises qui évitent de publier une perte.

Tableau 187 : Calcul du montant moyen par moyenne pondérée en 2001, 2002 et 2004

Année 2001													total	
De	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,01	0,011		
À	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,01	0,011	0,012		
Observé	5	8	11	9	9	6	9	13	10	7	7	7		
Ajusté	4,25	4,39	4,53	4,68	4,84	5	5,17	5,35	5,54	5,74	5,94	6,16		
Surnombre	0,75	3,61	6,47	4,32	4,16	1	3,83	7,65	4,46	1,26	1,06	0,84	39,41	
montant manipulé	0,0005	0,0015	0,0025	0,0035	0,0045	0,0055	0,0065	0,0075	0,0085	0,0095	0,0105	0,0115		
montant x surnombre	0,00037	0,00541	0,01617	0,01512	0,01872	0,0055	0,02489	0,05737	0,03791	0,01197	0,01113	0,00966	0,214245	(bénéfices irrégulièrement publiés)
moyenne pondérée													0,005436	(bénéfice moyen irrégulièrement publié)
De	-0,001	-0,009	-0,01	-0,011	-0,019	-0,02	-0,021	-0,029	-0,03	-0,031	-0,035			
à	0	-0,001	-0,009	-0,01	-0,011	-0,019	-0,02	-0,021	-0,029	-0,03	-0,031			
Observé	2	16	4	0	12	1	1	13	0	3	2			
Ajusté	4,12	28,93	3,17	3,09	21,98	2,45	2,39	17,18	1,93	1,89	7,16			
manquant	-2,12	-12,93	0,83	-3,09	-9,98	-1,45	-1,39	-4,18	-1,93	1,11	-5,16		-40,29	
manquant corrigé	-2,12	-12,93	0,83	-3,09	-9,98	-1,45	-1,39	-4,18	-1,93	1,11	-4,28		-39,41	
montant manipulé	-0,0005	-0,005	-0,0095	-0,0105	-0,015	-0,0195	-0,0205	-0,025	-0,0295	-0,0305	-0,033			
montant x nombre d'entreprises	0,00106	0,06465	-0,00789	0,03244	0,1497	0,02827	0,02849	0,1045	0,05693	-0,03386	0,14124		0,56556	(déficit irrégulièrement évité)
moyenne pondérée													0,014351	(déficit moyen irrégulièrement évité)
total moyen													**0,019787**	**Soit 1,98% de l'AT**

Année 2002											total	
De	0	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009		
à	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,01		
Observé	6	10	10	8	8	7	12	11	12	12		
Ajusté	4,52	4,67	4,82	4,97	5,14	5,31	5,49	5,68	5,87	6,08		
Surnombre	1,48	5,33	5,18	3,03	2,86	1,69	6,51	5,32	6,13	5,92	43,45	
montant manipulé	0,0005	0,0015	0,0025	0,0035	0,0045	0,0055	0,0065	0,0075	0,0085	0,0095		
montant x surnombre	0,0007	0,008	0,013	0,0106	0,0129	0,0093	0,0423	0,0399	0,05211	0,0562	0,245	(bénéfices irrégulièrement publiés)
moyenne pondérée											0,00564	(bénéfice moyen irrégulièrement publié)

De	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,035	-0,04	-0,041		
à	0	-0,005	-0,01	-0,015	-0,02	-0,025	-0,03	-0,035	-0,04		
Observé	11	13	12	12	4	7	4	4	0		
Ajusté	20,68	17,89	15,6	13,7	12,11	10,77	9,63	8,65	1,63		
manquant	-9,68	-4,89	-3,6	-1,7	-8,11	-3,77	-5,63	-4,65	-1,63	-43,66000	
manquant corrigé	-9,68	-4,89	-3,6	-1,7	-8,11	-3,77	-5,63	-4,65	-1,42	-43,45000	
montant manipulé	-0,003	-0,008	-0,013	-0,018	-0,023	-0,028	-0,033	-0,038	-0,0405		
montant x nombre d'entreprises	0,0242	0,0367	0,045	0,0298	0,1825	0,1037	0,183	0,1744	0,05751	0,83664	(déficit irrégulièrement évité)
moyenne pondérée										-0,01926	(déficit moyen irrégulièrement évité)
total moyen										**0,02489**	**Soit 2,49% de l'AT**

Année 2004								total	
De	0	0,001	0,002	0,005	0,01	0,015	0,016		
à	0,001	0,002	0,005	0,01	0,015	0,016	0,017		
Observé	3	7	16	29	35	8	8		
Ajusté	3,31	3,41	10,93	20,83	24,85	5,55	5,77		
Surnombre	-0,31	3,59	5,07	8,17	10,15	2,45	2,23	31,35	
montant manipulé	0,0005	0,0015	0,0035	0,0075	0,0125	0,0155	0,0165		
montant x surnombre	-0,00016	0,00538	0,01774	0,06127	0,12687	0,03797	0,03679	0,285895	(bénéfices irrégulièrement publiés)
moyenne pondérée								0,009119	(bénéfice moyen irrégulièrement publié)

De	-0,001	-0,002	-0,003	-0,005	-0,01	-0,02	-0,03		
à	0	-0,001	-0,002	-0,003	-0,005	-0,01	-0,02		
Observé	0	2	1	0	9	14	8		
Ajusté	3,21	3,11	3,02	5,77	13,04	21,3	16,55		
manquant	-3,21	-1,11	-2,02	-5,77	-4,04	-7,3	-8,55	-32	
manquant corrigé	-3,21	-1,11	-2,02	-5,77	-4,04	-7,3	-7,9	-31,35	
montant manipulé	-0,0005	-0,0015	-0,0025	-0,004	-0,0075	-0,015	-0,025		
montant x nombre d'entreprises	0,00160	0,00166	0,00505	0,02308	0,0303	0,1095	0,1975	0,3687	(déficit irrégulièrement évité)
moyenne pondérée								-0,01176	(déficit moyen irrégulièrement évité)
Total moyen								**0,02088**	**Soit 2,09% de l'AT**

BIBLIOGRAPHIE ET INDEX DES FIGURES ET TABLEAUX

Bibliographie

Ayers B. C., Jiang J. X. et Yeung P. E. (2006). Discretionary Acruals and Earnings Management: An Analysis of Pseudo Earnings Targets. *The Accounting Review* 81 (3): 617-652.

Ball R. et Brown P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research* 6 (2): 159-178.

Barth M. E., Nelson K. K. et Cram D. P. (2001). Accruals and the Prediction of Future Cash Flows. *Accounting Review* 76 (1): 27-58.

Bartov E., Givoly D. et Hayn C. (2002). The rewards to meeting or beating earnings expectations. *Journal of Accounting & Economics* 33 (2): 173-204.

Barua A., Legoria J. et Moffitt J. S. (2006). Accruals Management to Achieve Earnings Benchmarks: A Comparison of Pre-managed Profit and Loss Firms. *Journal of Business Finance & Accounting* 33 (5/6): 653-670.

Beaver W. H. (1968). The Information Content of Annual Earnings Announcements. *Journal of Accounting Research* 6 (3): 67-92.

Beaver W. H., McNichols M. F. et Nelson K. K. (2007). An Alternative Interpretation of the Discontinuity in Earnings Distributions. *Review of Accounting Studies* 12 (4): 525-556.

Beneish M. D. (1997). Detecting GAAP Violation : Implications for Assessing Earnings Management among Firms with Extreme Financial Performance. *Journal of Accounting and Public Policy* 16 (3): 271-309.

Beneish M. D. (1999). The Detection of Earnings Manipulation. *Financial Analysts' Journal* 55 (5): 24-36.

Bhattacharya U., Daouk H. et Welker M. (2003). The World Price of Earnings Opacity. *Accounting Review* 78 (3): 641-678.

Bisson B., Dumontier P. et Janin R. (2004), Les entreprises non cotées manipulent-elles leurs résultats ?, 3ème colloque international : gouvernance et juricomptabilité, Montréal.

Breton G. et Stolowy H. (2004). Accounts Manipulation: A Literature Review and Proposed Conceptual Framework. *Review of Accounting and Finance* 3 (1): 5-66.

Brown L. D. (1996). Analyst Forecasting Errors and Their Implications for Security Analysis: An Alternative Perspective. *Financial Analysts Journal* 52 (1): 40-47.

Brown L. D. (1997). Analyst Forecasting Errors: Additional Evidence. *Financial Analysts Journal* 53 (6): 81-88.

Brown L. D. (2001). A Temporal Analysis of Earnings Surprises: Profit versus Losses. *Journal of Accounting Research* 39 (2): 221-241.

Brown L. D. et Higgins H. N. (2001). Managing Earnings Surprises in the US versus 12 other Countries. *Journal of Accounting and Public Policy* 20 (4-5): 373-398.

Brown L. D. et Caylor M. L. (2005). A Temporal Analysis of Earnings Management Thresholds: Propensities and Valuation Consequences. *Accounting Review* 80 (2): 423-440.

Burgstahler D. et Dichev I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting & Economics* 24 (1): 99.

Burgstahler D. et Eames M. (2006). Management of Earnings and Analysts' Forecasts to Achieve Zero and Small Positive Earnings Surprises. *Journal of Business Finance and Accounting* 33 (5): 633-652.

Burgstahler D., Hail L. et Leuz C. (2006). The Importance of Reporting Incentives: Earnings Management in European Private and Public Firms. *The Accounting Review* 81 (5): 983-1016.

Burgstahler D. C. et Eames M. J. (2003). Earnings Management to Avoid Losses and Earnings Decrease: Are Analysts Fooled? *Contemporary Accounting Research* 20 (2): 253-294.

Burgstahler D. C., Elliott W. B. et Hanlon M. (2003). How Firms Avoid Losses: Evidence of Use of the Net Deferred Tax Asset Account. Working Paper. University of Michigan Business School.

Caneghem T. V. (2006). Audit Quality, Materiality & Threshold-Induced Earnings Management. *The ICFAI Journal of Audit Practice* 3 (2): 60-77.

Caramanis C. et Lennox C. (2008). Audit effort and earnings management. *Journal of Accounting & Economics* 45 (1): 116-138.

Carslaw C. (1988). Anomalies in accounting numbers : evidence of goal oriented behaviour". *The Accounting Review* 63 (2): 321-327.

Chalayer-Rouchon S., Degeorge F. et Le Nadant A.-L. (2001). Objectifs de la gestion des résultats et des marchés financiers. Dans *Faire de la recherche en comptabiltié financière*. 235-249.

Chambost I. (2007). *Contribution à l'analyse de la formation du jugement des analystes financiers sell-side*. Doctorat en sciences de gestion, Paris: Conservatoire National des Arts et Métiers.

Chen S., Lin B.-X., Wang Y. et Wu L. (2005). Detecting the Frequency and Magnitude of Earnings Management ; A Parametric Model and Empirical Analysis. Working Paper. University of Rhode Island.

Coase R. (1937). The Nature of the Firm. *Economica New Series* 4: 386-405.

Cohen D. A. (2003). Quality of Financial Reporting Choice: Determinants and Economic Consequences. Working Paper. Northwestern University.

Coppens L. et Peek E. (2005). An analysis of earnings management by European private firms. *Journal of International Accounting, Auditing and Taxation* 14 (1): 1-17.

Cornsweet T. C. (1974). *Visual Perception*. New York: Academic Press.

Cotter J., Tuna I. et Wysocki P. D. (2006). Expectations Management and Beatable Targets: How Do Analysts React to Explicit Earnings Guidance? *Contemporary Accounting Research* 23 (3): 593-624.

Das S. et Zhang H. (2003). Rounding-up in reported EPS, behavioral thresholds, and earnings management. *Journal of Accounting and Economics* 35 (1): 31-50.

Daske H., Gebhardt G. et McLeay S. (2006). The distribution of earning relative to targets in the European Union. *Accounting & Business Research* 36 (3): 137-168.

Davidson R., Goodwin-Stewart J. et Kent P. (2005). Internal governance structures and earnings management. *Accounting & Finance* 45 (2): 241-267.

DeAngelo L. (1986). Accounting Numbers as Market Valuation Substitutes: a Study of Management Buyouts of Public Stockholders. *The Accounting Review* 61: 400-420.

Dechow P. M. et Sloan R. G. (1991). Executive incentives and the horizon problem : An empirical investigation. *Journal of Accounting and Economics* 14 (1): 51-89.

Dechow P. M., Sloan R. G. et Sweeney A. P. (1995). Detecting Earnings Management. *Accounting Review* 70 (2): 193-225.

Dechow P. M. et Skinner D. J. (2000). Earnings management: reconciling the views of accounting academics, practitioners and regulators. *Accounting Horizons* 14 (2): 235-250.

Dechow P. M. et Dichev I. D. (2002). The Quality of Accruals and Earnings: The Role of Accrual Estimation Errors. *Accounting Review* 77 (4): 35-60.

Dechow P. M., Richardson S. A. et Tuna I. (2003). Why Are Earnings Kinky? An Examination of the Earnings Management Explanation. *Review of Accounting Studies* 8 (2/3): 355-384.

Defond M. L. et Park C. W. (2001). The Reversal of Abnormal Accruals and the Market Valuation of Earnings Surprises. *Accounting Review* 76 (3): 375.

Degeorge F., Patel J. et Zeckhauser R. (1999). Earnings Management to Exceed Thresholds. *The Journal of Business* 72 (1): 1-33.

Dichev I. D. et Skinner D. J. (2002). Large Sample Evidence on the Debt Covenant Hypothesis. *Journal of Accounting Research* 40 (4): 1091-1123.

Dreman D. N. et Berry M. A. (1995). Analyst Forecasting Errors and Their Implications for Security Analysis. *Financial Analysts Journal* 51 (3): 30.

Dumontier P. et Labelle R. a. (1998). Accounting earnings and firm valuation: the French case. *European Accounting Review* 7 (2): 163-183.

Dumontier P. et Raffournier B. (1999). Vingt ans de recherche positive en comptabilité financière. *Comptabilité Contrôle Audit* (Les vingt ans de l'AFC): 179-197.

Durtschi C. et Easton P. (2005). Earnings Management ? The shapes of the Frequency Distributions of Earnings Metrics Are Not Evidence Ipso Facto. *Journal of Accounting Research* 43 (4): 557-593.

Frank M. M. et Rego S. O. (2006). Do Managers use the Valuation Allowance Account to Manage Earnings Around Certain Earnings Targets ? *Journal of the American Taxation Association* 28 (1): 43-65.

Freedman D. et Diaconis P. (1981). On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57: 453-476.

Fudenberg D. et Tirole J. (1995). A theory of income and dividend smoothing based on incumbency rents. *Journal of Political Economy* 103 (1): 75-93.

Givoly D. et Hayn C. (2002). Rising Conservatism: Implications for Financial Analysis. *Financial Analysts Journal* 58 (1): 56-74.

Glass A. L. et Holyoak K. J. (1986). *Cognition*. New York, Random House.

Glaum M., Lichtblau K. et Lindemann J. (2004). The Extent of Earnings Management in the US & Germany. *Journal of International Accounting Research* 3 (2): 45-77.

Goncharov I. et Zimmermann J. (2006). Earnings Management when Incentives Compete: The Role of Tax Accounting in Russia. *Journal of International Accounting Research* 5 (1): 41-65.

Graham J. R., Harvey C. R. et Rajgopal S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting & Economics* 40 (1-3): 3-73.

Graham J. R., Harvey C. R. et Rajgopal S. (2006). Value Destruction and Financial Reporting Decisions. *Financial Analysts Journal* 62 (6): 27-39.

Halzen F. et Martin A. D. (1984). *Quarks and Leptons: An introductory course in modern particle physics.* New York: NY: Wiley.

Hayn C. (1995). The information content of losses. *Journal of Accounting and Economics* 20 (2): 125-153.

Healy P. (1985). The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7 (1-3): 85-107.

Healy P. M. et Whalen J. M. (1999). A Review of the Earnings Management Literature and its Implications for Standard Setting. *Accounting Horizons* 13 (4): 365-383.

Holland D. et Ramsey A. (2003). Do Australian companies manage earnings to meet simple earnings benchmarks ? *Accounting and Finance* 43: 41-62.

Holland K. et Jackson R. H. G. (2004). Earnings management and deferred tax. *Accounting & Business Research* 34 (2): 101-123.

Jeanjean T. (2001). Incitations et contraintes à la gestion du résultat. *Comptabilité Contrôle Audit* 7 (1): 61-76.

Jeanjean T. (2002). *Gestion du résultat et gouvernement d'entreprise : étude des déterminants et formulation d'un modèle de mesure*. Doctorat en sciences de gestion, Paris: Université Paris IX Dauphine.

Jensen M. C. et Meckling W. (1976). Theory of the firm : managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305-360.

Jones J. J. (1991). Earnings Management During Import Relief Investigations. *Journal of Accounting Research* 29 (2): 193-228.

Kahneman D. et Tversky A. (1979). Prospect Theory : an Analysis of Decision under Risk. *Econometrica* 47 (2): 263-291.

Kasznik R. (1999). On the Association between Voluntary Disclosure and Earnings Management. *Journal of Accounting Research* 37 (1): 57-81.

Kasznik R. et McNichols M. F. (2002). Does Meeting Earnings Expectations Matter? Evidence from Analyst Forecast Revisions and Share Prices. *Journal of Accounting Research* 40 (3): 727-759.

Kinney W., Burgstahler D. et Martin R. (2002). Earnings Surprise "Materiality" as Measured by Stock Returns. *Journal of Accounting Research* 40 (5): 1297-1329.

Leuz C., Nanda D. et Wysocki P. D. (2003). Earnings Management and Investor Protection: An International Comparison. *Journal of Financial Economics* 69 (3): 505-527.

Levitt A. L. (1998), Remarks by Chairman Arthur Levitt Securities and Exchange Commission, The "Numbers Game." NYU Center for Law and Business, New York.

Lopez T. J. et Rees L. (2002). The Effect of Beating and Missing Analysts' Forecasts in the Information Content of Unexpected Earnings. *Journal of Accounting, Auditing & Finance* 17 (2): 155-184.

Mard Y. (2002). *Etude des pratiques de gestion de résultats comptables : application empirique sur un échantillon d'entreprises françaises cotées*. Doctorat en sciences de gestion, Paris: Université Paris Dauphine.

Mard Y. (2004). Les sociétés françaises cotées gèrent-elles leurs chiffres comptables afin d'éviter les pertes et les baisses de résultats ? *Comptabilité Contrôle Audit* 10 (2): 73-98.

Matsumoto D. A. (2002). Management's Incentives to Avoid Negative Earnings Surprises. *Accounting Review* 77 (3): 483-514.

McGough R. (2000). How 'Round-Ups' Can Give Stocks a Hard Ride. *Wall Street Journal - Eastern Edition* 236 (14): C1.

McNichols M. et Wilson P. (1988). Evidence of Earnings Management from the Provision for Bad Debts. *Journal of Accounting Research* 26 (supplement): 1-31.

McNichols M. F. (2000). Research design issues in earnings management studies. *Journal of Accounting & Public Policy* 19 (4/5): 313-345.

McNichols M. F. (2003). Discussion of "Why are Earnings Kinky? An Examination of the Earnings Management Explanation". *Review of Accounting Studies* 8 (2-3): 385-391.

McVay S. E., Nagar V. et Tang V. W. (2006). Trading Incentives to Meet the Analyst Forecast. *Review of Accounting Studies*:

Payne J. L. et Robb S. W. G. (2000). Earnings Management: The Effect of Ex Ante Earnings Expectations. *Journal of Accounting, Auditing & Finance* 15 (4): 371-392.

Payne J. L. et Thomas W. B. (2004). Implications of Unexpectedly Meeting and Missing Earnings Thresholds on the Returns/Earnings Relation. Working Paper. University of Kentucky.

Petrovits C. M. (2006). Corporate-sponsored foundations and earnings management. *Journal of Accounting and Economics* 41 (3): 335-362.

Phillips J., Pincus M. et Rego S. O. (2003). Earnings Management: New Evidence Based on Deferred Tax Expense. *The Accounting Review* 78 (2): 491-521.

Prat dit Hauret C. (2007). Ethique et décisions d'audit. *Comptabilité Contrôle Audit* 13 (1): 69-85.

Raffournier B. (1990). La théorie "positive" de la comptabilité : une revue de la littérature. *Economie et Sociétés : série sciences de gestion* 16: 137-166.

Roychowdhury S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics* 42 (3): 335-370.

Schrand C. M. et Wong M. H. F. (2003). Earnings Management Using the Valuation Allowance for Deferred Tax Assets under SFAS No. 109. *Contemporary Accounting Research* 20 (3): 579-611.

Scott D. W. (1979). On optimal and data-based histograms. *Biometrika* 66 (3): 605-610.

Scott D. W. (1992). *Multivariate Density Estimation*. New York: Wiley: New York: Willey.

Shen C.-H. et Chih H.-L. (2005). Investor protection, prospect theory, and earnings management: An international comparison of the banking industry. *Journal of Banking & Finance* 29 (10): 2675-2697.

Shuto A. (2008). Earnings Management to Avoid Earnings Decreases: A Comparative Analysis of Consolidated Earnings and Parent-Only Earnings. Working Paper. Kobe University.

Silverman B. W. (1986). *Density Estimation for Statistics and Data Analysis*. CRC Press.

Simon H. A. (1947). *Administrative Behavior*. New York: New York: Macmillan.

Skinner D. J. et Sloan R. G. (2002). Earnings Surprises, Growth Expectations, and Stock Returns or Don't Let an Earnings Torpedo Sink Your Portfolio. *Review of Accounting Studies* 7 (2/3): 289-312.

Takeuchi Y. (2004). On a statistical method to detect discontinuity in the distribution function of reported earnings. *Mathematics and computers in simulation* 64 (1): 103-111.

Thomas J. K. (1989). Unusual Patterns in Reported Earnings. *Accounting Review* 64 (4): 773.

Vidal O. (2007), La gestion du résultat pour atteindre des seuils, Congrès de l'Association Francophone de Comptabilité, Poitiers, France.

Vidal O. (2007), Le rôle des aversions psychologiques dans l'évitement des seuils comptables : une étude exploratoire, Conférence internationale de Gouvernance d'Entreprise, Genève, Suisse.

Wasley C. E. et Wu J. S. (2005). Why Do Managers Voluntarily Issue Cash Flow Forecasts? *Journal of Accounting Research* 44 (2): 389-.

Watts R. L. et Zimmerman J. L. (1978). Towards a Positive Theory of the Determination of Accounting Standards. *The Accounting Review* 53: 112-134.

Watts R. L. et Zimmerman J. L. (1986). *Positive Accounting Theory*. London: Prentice-Hall international.

Williamson O. E. (1985). *The Economic Institutions of Capitalism*. New York / London: UK : Free Press.

Xue Y. (2004). Information Content of Earnings Management: Evidence from Managing Earnings to Exceed Thresholds. Working Paper. George Washington University.

Yu Q., Du B. et Sun Q. (2006). Earnings Management at rights issues thresholds - Evidence from China. *Journal of Banking & Finance* 30 (12): 3453-3468.

Index des figures

Figure 1 : Étude exploratoire. Constatation des irrégularités de distribution des résultats de 1999 à 2002 par les entreprises du SBF 250 11
Figure 2 : La question de recherche dans le cadre plus général de la recherche en comptabilité sur la valeur des informations comptables 13
Figure 3 : La qualité des informations comptables 16
Figure 4 : Le filtre de l'information comptable....18
Figure 5 : Les deux types de manipulations comptables 20
Figure 6 : Les trois démarches statistiques d'étude de la gestion du résultat 24
Figure 7 : La double discontinuité 33
Figure 8 : Interprétation de la double discontinuité....34
Figure 9 : Les actifs financiers comme explication à l'irrégularité 37
Figure 10 : De l'observation des discontinuités à l'évitement des seuils....68
Figure 11: Absence d'effet de seuil 71
Figure 12 : Effet de seuil stricto sensu 71
Figure 13 : Comparaison des distributions des entreprises cotées américaines et allemandes (Glaum, Lichtblau et Lindemann, 2004) : mise en évidence de deux irrégularités distinctes 74
Figure 14 : Les facteurs psychologiques comme variables modératrices : effet sur le dirigeant....78
Figure 15 : Les facteurs psychologiques comme variables modératrices : effet sur les investisseurs....79
Figure 16 : Les motivations à la gestion du résultat pour éviter les seuils : un cadre d'analyse 80
Figure 17 : Distribution du résultat mis à l'échelle par l'Actif Total de (n-1) des entreprises du SBF250 en 1997 et 2002 87
Figure 18 : Figure issue de l'article de Glaum, Lichtblau et Lindemann (2004). La population représente près de 4 000 entreprises américaines sur 10 ans, soit 38 714 observations 88
Figure 19 : Figure issue de l'article de Burgstahler et Dichev, 1997 89
Figure 20 : Figure issue de l'article de Dechow, Richardson et Tuna, 2003 90
Figure 21 : Mesure par moyenne arithmétique des deux intervalles adjacents 92
Figure 22 : Moyenne arithmétique, non monotonie et non linéarité de la distribution 93
Figure 23 : Utilisation d'un intervalle potentiellement irrégulier pour mesurer l'irrégularité94
Figure 24 : Influence de la largeur de l'intervalle d'étude sur les mesures par moyenne arithmétique....95
Figure 25 : Mesure par interpolation linéaire 96
Figure 26 : Mesure par extrapolation 97
Figure 27 : Mesure par symétrie 98
Figure 28 : Mesure par interpolation exponentielle....100
Figure 29 : Prise en compte des transferts d'observations....101
Figure 30 : Limite du calcul en deux temps de prise en compte des transferts d'observations 102

Figure 31 : Indicateur d'asymétrie d/g et non linéarité de la distribution 105
Figure 32 : Indicateur d'asymétrie et taille de l'échantillon 105
Figure 33 : Comparaison des deux indicateurs d'asymétrie 107
Figure 34 : Indicateur d'asymétrie (r-l)/(r+l) et non linéarité de la distribution 107
Figure 35 : Choix de la mesure appropriée en fonction de la position du seuil sur la distribution 112
Figure 36 : Remise en cause de l'allure en cloche de la distribution 114
Figure 37 : Évolution de moyenne, médiane, et écart-type de la variable IB / (AT-IB) 132
Figure 38 : Évolution de max et min de la variable IB / (AT-IB) 133
Figure 39 : Observations nécessaires à la mesure par symétrie 139
Figure 40 : Illustration des quatre séries de mesures par interpolation réalisées.....141
Figure 41 : Illustration graphique des interpolations linéaires menées sur la seule année 1992 142
Figure 42 : Illustration graphique des interpolations logarithmiques menées sur la seule année 1992.....143
Figure 43 : Illustration graphique des interpolations exponentielles menées sur la seule année 1992.....144
Figure 44 : Représentation graphique des irrégularités mesurées sur [-0,005 ; 0[.....151
Figure 45 : Représentation graphique des irrégularités droites mesurées sur [0 ; 0,005[.....153
Figure 46 : Classification hiérarchique des mesures non paramétriques de l'irrégularité gauche.....156
Figure 47 : Classification hiérarchique des mesures non paramétriques de l'irrégularité droite 157
Figure 48 : Comparaison des tendances des mesures non paramétriques (irrégularité gauche) 158
Figure 49 : Comparaison des tendances des mesures non paramétriques (irrégularité droite) 160
Figure 50 : Choix de la mesure appropriée en fonction de la position du seuil sur la distribution 161
Figure 51 : Position du seuil et choix de la mesure appropriée en 1992.....162
Figure 52 : évolution de l'irrégularité (avec courbe de tendance) comparée aux autres mesures.....164
Figure 53 : Comparaison graphique de l'évolution des irrégularités (en valeur absolue) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire.....168
Figure 54 : Comparaison graphique de l'évolution des irrégularités (tenant compte du signe de l'irrégularité) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire.....169
Figure 55 : évolution de l'indicateur d'asymétrie d/g entre 1992 et 2004 171
Figure 56 : Comparaison graphique des indicateurs d'asymétrie d/g et (d-g)/(d+g) en faisant varier la largeur des intervalles.....173
Figure 57 : Comparaison graphique des indicateurs d'asymétrie d/g et (d-g)/(d+g) avec courbe de tendance 174
Figure 58 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles faibles ; 0,001 et 0,005) 178
Figure 59 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles intermédiaires ; 0,010 et 0,0122) 180

Figure 60 : Distribution des résultats des entreprises françaises cotées en 1997 (représentations avec largeurs d'intervalles importantes ; 0,034 et 0,050) 181
Figure 61 : Contrainte de largeur maximale dans le calcul par moyenne 182
Figure 62 : Contrainte de largeur maximale dans le calcul par symétrie 184
Figure 63 : Mesure par symétrie de l'irrégularité de distribution autour du résultat nul ; entreprises françaises cotées du SBF 250 en 1994 et en 1998 185
Figure 64 : Observation des 4 intervalles encadrant le seuil pour mesurer les irrégularités droite et gauche par moyenne 186
Figure 65 : Irrégularités (mesurées par moyenne en 1997) en fonction de la largeur d'intervalle 186
Figure 66 : Lorsque la largeur est très importante, l'irrégularité tend vers l'infini car l'effectif attendu (théorique) tend vers zéro 187
Figure 67 : Irrégularités relatives (mesurées par moyenne) en fonction de la largeur : zoom sur les largeurs pertinentes 188
Figure 68 : Irrégularités (mesurées par symétrie en 1997) en fonction de la largeur d'intervalle 189
Figure 69 : Irrégularités relatives (en 1997) mesurée par symétrie : zoom sur les largeurs pertinentes.....190
Figure 70 : Comparaison de l'influence de la largeur d'intervalle sur les mesures par symétrie et par moyenne ; irrégularité droite 191
Figure 71 : Comparaison de l'influence de la largeur d'intervalle sur les mesures par symétrie et par moyenne ; irrégularité gauche 192
Figure 72 : Exemple de distribution.....193
Figure 73 : Exemples d'irrégularités.....193
Figure 74 : Représentation graphique de la zone d'irrégularité potentielle associée à une distribution 194
Figure 75 : Représentation graphique de la significativité des irrégularités 194
Figure 76 : Largeur et hauteur de l'irrégularité, deux mesures complémentaires 195
Figure 77 : Distribution de la variable résultat non mise à l'échelle en 1992.....200
Figure 78 : Distribution de la variable résultat (non mise à l'échelle) agrégée sur 13 ans 201
Figure 79 : Allure apparente de la distribution des résultats non mis à l'échelle 202
Figure 80 : Comparaison graphique des distributions selon les deux transformations de la variable AT ; « zoom » sur la partie centrale des distributions.....206
Figure 81 : Diagrammes en boites à moustaches des deux transformations de la variable AT.....207
Figure 82 : Distribution des trois variables de mise à l'échelle 217
Figure 83 : Comparaison des entreprises identifiées comme faiblement bénéficiaires/déficitaires en utilisant les variables CA et AT 221
Figure 84 : Identification, en utilisant les variables CA et VM, des petites performances identifiées en utilisant la variable AT 223
Figure 85 : Identification, en utilisant les variables AT et VM, des petites performances identifiées en utilisant la variable CA.....224
Figure 86 : Identification, en utilisant les variables AT et CA, des petites performances identifiées en utilisant la variable VM 225
Figure 87 : Représentation graphique des irrégularités ; résultat mis à l'échelle par valeur de marché.....228

Figure 88 : Comparaison des irrégularités mesurées en mettant à l'échelle par les différents dénominateurs 229
Figure 89 : Distribution agrégée sur 13 ans du résultat mis à l'échelle par le chiffre d'affaires (partie centrale de la distribution).....231
Figure 90 : Exemple de distribution lisse irrégulière 239
Figure 91 : Distribution des résultats non réduits 242
Figure 92 : Distribution des résultats mis à l'échelle par la variable (AT-IB) 244
Figure 93 : Représentation schématique des différences entre distribution empirique et distribution Normale.....245
Figure 94 : Transformation de la zone irrégulière lors de l'agrégation des données ; comparaison entre agrégation des données brutes et centrées-réduites 255
Figure 95 : Exemples de lois de distribution symétriques.....257
Figure 96 : Exemples de lois de distribution dissymétriques 257
Figure 97 : Exemples de lois monotones 258
Figure 98 : Représentation graphique de la mesure des irrégularités par une méthode paramétrique 261
Figure 99 : évolution des indicateurs de tendance des distributions avant et après correction.....266
Figure 100 : Comparaison de la moyenne et de la médiane de chacune des distributions annuelles 267
Figure 101 : Identification du sommet des distributions entre 2001 et 2004 268
Figure 102 : Comparaison des sommets des distributions agrégées de 2001 à 2004 centrées sur moyenne et médiane 269
Figure 103 : Comparaison des sommets des distributions agrégées de 1992 à 2004 centrées sur moyenne et médiane 270
Figure 104 : Représentation graphique de l'ajustement de la distribution par une loi logistique 273
Figure 105 : Ajustement et représentation graphique de l'ajustement sur l'intervalle]-0,8 ; 0,8[.....274
Figure 106 : Distribution des écarts entre distribution empirique et loi ajustée sur l'intervalle]-0,8 ; 0,8[.....275
Figure 107 : « Q-Q plot » de l'ajustement Logistique 276
Figure 108 : Transposition de la demi-distribution des valeurs négatives.....277
Figure 109 : Les lois monotones sont privilégiées dans le cadre des ajustements partiels 279
Figure 110 : Caractéristiques des populations d'entreprises allemandes et britanniques 283
Figure 111 : Représentation graphique des distributions britanniques centrées-réduites ; 1999-2003 285
Figure 112 : Représentation graphique des distributions allemandes centrées-réduites ; 2001-2003 287
Figure 113 : Représentations graphiques de la distribution britannique centrée, réduite et agrégée 288
Figure 114 : Représentations graphiques de la distribution allemande centrée, réduite et agrégée 289
Figure 115 : Comparaison de la loi ajustée sur la seule partie centrale des données, avec la distribution globale 293

Figure 116 : Représentation graphique de la demi-distribution gauche des sociétés allemandes 299
Figure 117 : Comparaison graphique de la méthode d'estimation Pareto2 par rapport aux estimations non paramétriques 302
Figure 118 : Représentation graphique des irrégularités mesurées par l'ajustement Logistique 307
Figure 119 : Représentation graphique des irrégularités par ajustement Pareto2 311
Figure 120 : Comparaison des irrégularités mesurées par ajustement Logistique et Pareto2 312
Figure 121 : Comparaison des irrégularités mesurées par les méthodes paramétrique avec les irrégularités mesurées par des méthodes non paramétriques.....313
Figure 122 : Partie gauche de la distribution (centrée sur la médiane, réduite, et transposée) des résultats des entreprises française en 1997 317
Figure 123 : Différentes interprétations de la comparaison des irrégularités droite et gauche.....318
Figure 124 : Recherche de la largeur maximale de l'intervalle irrégulier.....319
Figure 125 : Pourquoi l'ajustement Logistique n'est pas retenu pour mesurer les irrégularités maximales 326
Figure 126 : Largeur maximale de l'intervalle irrégulier en 1992 327
Figure 127 : Comment identifier le nombre total d'entreprises manipulatrices ?.....329
Figure 128 : Représentation graphique des largeurs maximales à droite, à gauche, et leur différence.....330
Figure 129 : Courbes de tendance (logarithmique) de l'évolution des irrégularités maximales 332
Figure 130 : Illustration de la démarche de détermination du montant maximal manipulé 333
Figure 131 : représentation graphique de la part des entreprises déficitaires identifiées comme manipulant leurs comptes pour éviter le seuil du résultat nul, par rapport au total des entreprises irrégulièrement faiblement bénéficiaires 335
Figure 132 : Illustration de la détermination du montant moyen manipulé sous hypothèse de répartition homogène des irrégularités 336
Figure 133 : Principe de détermination du montant moyen manipulé par moyenne pondérée 337
Figure 134 : Évolution des montants moyens manipulés (calcul approximatif) par les entreprises évitant de publier une perte 340
Figure 135 : Comparaison de l'évolution des irrégularités lorsque le seuil de significativité est modifié 342
Figure 136 : Représentation graphique de la demi-distribution gauche des résultats (transposés) en 2002.....451
Figure 137 : Représentation graphique de la demi-distribution gauche des résultats (transposés) en 2003.....452

Index des tableaux

Tableau 1 : Typologie des méthodologies d'étude de la gestion du résultat.....23
Tableau 2 : Typologie des accruals.....25
Tableau 3 : Typologie des seuils et conséquences méthodologiques 69
Tableau 4 : Liens entre effet de seuil stricto sensu et effet d'évitement.....72
Tableau 5 : Mesures non paramétriques des irrégularités. Postulats implicites et synthèse des limites et avantages 109
Tableau 6 : Synthèse des limites et avantages des indicateurs d'asymétrie 110
Tableau 7 : Comparaison des caractéristiques mathématiques des distributions des deux transformations de la variable AT utilisées comme indicateur de taille.....120
Tableau 8 : Comparaison de la part du résultat respectivement dans les variables AT, le CA et la VM 122
Tableau 9 : Base constituée sur les seules sociétés actives en 2005 (biais du survivant) 125
Tableau 10 : Base constituée sur les sociétés actives et inactives de 1992 à 2005 125
Tableau 11 : Synthèse des caractéristiques de la base de données (entreprises françaises) avant ajustement. 128
Tableau 12 : Composition de la base après corrections ; variable IB / (AT-IB).....131
Tableau 13 : Caractéristiques de la base après corrections ; variable IB / (AT-IB).....132
Tableau 14 : Composition sectorielle de l'échantillon.....134
Tableau 15 : Mesures non paramétriques des irrégularités. Classement par méthodologie mathématique. 136
Tableau 16 : Caractéristiques de la distribution agrégée des 13 années, et largeur théorique de l'intervalle d'étude 137
Tableau 17 : Caractéristiques des 13 distributions, et largeur théorique de l'intervalle d'étude.....137
Tableau 18 : Tableau récapitulatif des effectifs autour du seuil résultat nul.....138
Tableau 19 : Tableau récapitulatif des effectifs autour de médiane x 2 et Moyenne x 2 .140
Tableau 20 : Équations des régressions permettant d'estimer les effectifs par interpolation pour la seule année 1992 145
Tableau 21 : Synthèse des R^2 des (9 x 13 =) 117 régressions.....145
Tableau 22 : Tableau des effectifs attendus par interpolation 146
Tableau 23 : Intervalles de référence pour les mesures par extrapolation.....147
Tableau 24 : Synthèse des extrapolations sur 4 intervalles.....148
Tableau 25 : Synthèse des extrapolations sur 8 intervalles.....148
Tableau 26 : Estimations selon les mesures par extrapolation.....149
Tableau 27 : Mesures non paramétriques de l'irrégularité sur [-0,005 ; 0[.....150
Tableau 28 : Mesures non paramétriques de l'irrégularité droite sur [0 ; 0,005[.....152
Tableau 29 : Volatilité comparée des différentes mesures non paramétriques.....154
Tableau 30 : Qualité comparée des différentes régressions.....155
Tableau 31 : Comparaison des coefficients alpha (pente des régressions) ; irrégularité gauche.....159

Tableau 32 : Comparaison des coefficients alpha (pente des régressions) ; irrégularité droite 160
Tableau 33 : Emplacement de l'irrégularité dans la distribution annuelle et choix de la méthode appropriée (hypothèse sous-jacente : la distribution régulière a une forme de cloche) 163
Tableau 34 : Série de mesures optimale si la distribution attendue est en forme de cloche 163
Tableau 35 : Mesures de l'irrégularité (relative) à droite et à gauche du seuil résultat nul ; symétrie par rapport à la moyenne, et interpolation linéaire 167
Tableau 36 : Calcul des indicateurs d'asymétrie d/g et (d-g)/(d+g) ; entreprises françaises cotées de 1992 à 2004 172
Tableau 37 : Rappel des paramètres des distributions annuelles de IB/(AT-IB) 178
Tableau 38 : Largeur maximale année par année dans le cas de l'étude des entreprises françaises cotées ; calcul par moyenne arithmétique 183
Tableau 39 : Largeur maximale année par année dans le cas de l'étude des entreprises françaises cotées ; calcul par symétrie 185
Tableau 40 : paramètres des distributions du résultat net non mis à l'échelle 199
Tableau 41 : Comparaison du nombre d'observations de la variable AT et de la variable IB dans la base 204
Tableau 42 : Comparaison des distributions des deux transformations de la variable AT 205
Tableau 43 : Comparaison des deux transformations de la variable AT sur l'intervalle [-1 ;1[.....205
Tableau 44 : Comparaison des deux transformations de la variable AT sur l'intervalle [−0,5 ;0,5[.....206
Tableau 45 : Comparaison du nombre d'entreprises faiblement performantes d'après les variables IB/$AT_{(n-1)}$ et IB/(AT-IB) 208
Tableau 46 : Entreprises faiblement performantes identifiées par un seul des deux dénominateurs 209
Tableau 47 : Observation de la variable CA dans la base après correction, et comparaison avec la variable IB 211
Tableau 48 : Observation de la variable VM dans la base, et comparaison avec la variable IB 212
Tableau 49 : Nombre d'observations communes aux trois dénominateurs 212
Tableau 50 : Caractéristiques des distributions des variables de mise à l'échelle 213
Tableau 51 : Corrélations des variables de mise à l'échelle entre elles 214
Tableau 52 : Nombre d'entreprises faiblement déficitaire ou faiblement bénéficiaire selon un au moins des ratios 214
Tableau 53 : Comparaison des dénominateurs des entreprises faiblement déficitaires ou faiblement bénéficiaires 215
Tableau 54 : Corrélations des dénominateurs des entreprises identifiées comme faiblement bénéficiaires/déficitaires 215
Tableau 55 : Identification des entreprises faiblement bénéficiaires ou déficitaires 219
Tableau 56 : Concentration des distributions 219
Tableau 57 : Entreprises faiblement bénéficiaires ou déficitaires 220
Tableau 58 : Irrégularités annuelles mesurées en mettant à l'échelle par l'actif total 226
Tableau 59 : Irrégularités annuelles mesurées en mettant à l'échelle par le chiffre d'affaires 227

Tableau 60 : Calcul des irrégularités (résultat mis à l'échelle par la variable VM) 228
Tableau 61 : Distribution des entreprises dont le chiffre d'affaires est manquant 230
Tableau 62 : Distribution des entreprises dont la valeur de marché est manquante.....231
Tableau 63 : Catalogue des lois testées par le logiciel Best-Fit.....259
Tableau 64 : Rappel des paramètres des distributions annuelles de la variable IB/(AT-IB) 263
Tableau 65 : Examen des valeurs extrêmes de la distribution.....264
Tableau 66 : Suppression des 9 valeurs extrêmes de la distribution 264
Tableau 67 : Caractéristiques de la distribution après suppression des 9 valeurs extrêmes 265
Tableau 68 : Analyse des différences entre les distributions avant et après suppression des valeurs extrêmes.....265
Tableau 69 : Paramètres des distributions annuelles « brutes » (résultat mis à l'échelle par (AT-IB) non centré-réduit) 271
Tableau 70 : Paramètres des distributions annuelles centrées-réduites.....271
Tableau 71 : Paramètres de la distribution des résultats centrés-réduits agrégés sur 13 ans 271
Tableau 72 : Trois ajustements de la distribution globale 272
Tableau 73 : Nombre d'observations strictement égales à zéro 278
Tableau 74 : Ajustement de la demi-distribution « droite » (valeurs positives) 280
Tableau 75 : Ajustement de la demi-distribution « gauche » (valeurs négatives) 281
Tableau 76 : Caractéristiques comparées des résultats mis à l'échelle des entreprises allemandes, britanniques et françaises 283
Tableau 77 : Comparaison moyenne / médiane des distributions annuelles ; population britannique.....285
Tableau 78 : Comparaison moyenne / médiane des distributions annuelles ; population allemande 286
Tableau 79 : Paramètres de la distribution des résultats britanniques centrés, réduits et agrégés 288
Tableau 80 : Paramètres de la distribution des résultats allemands centrés, réduits et agrégés 289
Tableau 81 : Ajustement de la loi de distribution des résultats britanniques.....290
Tableau 82 : Meilleur ajustement de la partie centrale de la distribution des résultats britanniques (intervalle [-0,9 et 0,9]).....291
Tableau 83 : Ajustement de la distribution des résultats allemands 292
Tableau 84 : Ajustement de la partie centrale ([-0,6 ; 0,6]) de la distribution des résultats allemands 293
Tableau 85 : Ajustement de la loi de distribution des résultats britanniques ; demi-distribution droite 295
Tableau 86 : Ajustement de la loi de distribution des résultats britanniques ; demi-distribution gauche.....296
Tableau 87 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution droite 297
Tableau 88 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution droite après suppression des valeurs extrêmes 298
Tableau 89 : Ajustement de la loi de distribution des résultats allemands ; demi-distribution gauche.....300
Tableau 90 : Ajustement de la distribution 1992 par une loi Logistique – partie 1 sur 3..304

Tableau 91 : Irrégularités mesurées par l'ajustement Logistique.....307
Tableau 92 : Ajustement de la demi-distribution gauche 1992 par une loi Pareto2 309
Tableau 93 : Distribution des irrégularités en 1992 ; ajustement Pareto2.....310
Tableau 94 : Irrégularités mesurées en ajustement la distribution par une loi Pareto2...311
Tableau 95 : Corrélations entre les différentes mesures 314
Tableau 96 : Significativité des irrégularités annuelles 321
Tableau 97 : Irrégularité annuelle minimum pour être significative 324
Tableau 98 : Irrégularité gauche (en 1992) mesurée en élargissant l'intervalle d'étude .327
Tableau 99 : Irrégularité droite (en 1992) mesurée en élargissant l'intervalle d'étude 328
Tableau 100 : Largeur maximale permettant de constater une irrégularité significative (> 40 %) 328
Tableau 101 : Irrégularités annuelles maximales 330
Tableau 102 : Nombre total d'entreprises bénéficiaires manipulatrices comparé au nombre maximum d'entreprises déficitaires manquantes, montant maximal manipulé 334
Tableau 103 : Calcul approximatif du montant moyen annuel manipulé par les entreprises évitant une perte 336
Tableau 104 : Bénéfices excédentaires (moyenne pondérée) des soc. faiblement bénéficiaires ; 2001.....338
Tableau 105 : Pertes cachées (moyenne pondérée) des entreprises faiblement bénéficiaires en 2001.....338
Tableau 106 : Largeur maximale permettant de constater une irrégularité et mesure des irrégularités annuelles maximales (seuil de significativité de 30 %) 341
Tableau 107 : Réexamen des mesures d'irrégularité en faisant varier la largeur d'intervalle en 2003 343
Tableau 108 : Entreprises publiant un résultat strictement nul 358
Tableau 109 : Données manquantes identifiées comme des « trous » dans les séries de données.....360
Tableau 110 : Données manquantes identifiées par confrontation (IB manquant et AT présent) 364
Tableau 111 : Données manquantes identifiées par confrontation (AT manquant et IB présent) 366
Tableau 112 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 1999 ; Société Systran.....368
Tableau 113 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 1999 ; Société ASF.....368
Tableau 114 : Comparaison des AT lus dans les bases Global et Thomson ; année 1999 ; Société ASF.....369
Tableau 115 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 2000 369
Tableau 116 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; années 2001 et 2002 ; Société Coflexip 370
Tableau 117 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; années 2001 et 2002 ; Société Grande Paroisse 370
Tableau 118 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; années 2001 et 2002 ; Société Gameloft.....370
Tableau 119 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; années 2001 et 2002 ; Société Guyanor.....370
Tableau 120 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 2003 371
Tableau 121 : Valeurs extrêmes du ratio $IB/AT_{(n-1)}$; année 2004 371

Tableau 122 : Valeurs extrêmes du ratio IB/(AT-IB) 372
Tableau 123 : Chiffre d'affaires strictement nul ou négatif dans la base....389
Tableau 124 : Variations extrêmes du chiffre d'affaires....390
Tableau 125 : Sélection des neuf cas dont les variations extrêmes ne sont pas confirmées par la suite 391
Tableau 126 : Caractéristiques des distributions annuelles de IB mis à l'échelle par le chiffre d'affaires 393
Tableau 127 : Distributions annuelles et effectifs irréguliers de la variable résultat mise à l'échelle par le chiffre d'affaires....394
Tableau 128 : Calcul des irrégularités de distribution du résultat mis à l'échelle par le chiffre d'affaires 397
Tableau 129 : Caractéristiques des distributions du résultat mis à l'échelle par la valeur de marché 398
Tableau 130 : Distributions annuelles et effectifs irréguliers de la variable résultat mise à l'échelle par la variable VM 398
Tableau 131 : Calcul des irrégularités (résultat mis à l'échelle par la valeur de marché) 402
Tableau 132 : Valeur extrême de l'année 1995....403
Tableau 133 : Valeurs extrêmes de l'année 1996 403
Tableau 134 : Valeurs extrêmes de l'année 1997 404
Tableau 135 : Valeurs extrêmes de l'année 1998 404
Tableau 136 : Valeurs extrêmes de l'année 1999 405
Tableau 137 : Valeurs extrêmes de l'année 2000 405
Tableau 138 : Valeurs extrêmes de l'année 2001 406
Tableau 139 : Valeurs extrêmes de l'année 2002 407
Tableau 140 : Valeurs extrêmes de l'année 2002 ; Société SOFCO 408
Tableau 141 : Valeurs extrêmes de l'année 2002 ; société Phone Systemes & Network 408
Tableau 142 : Valeurs extrêmes de l'année 2002 ; société Aurea....409
Tableau 143 : Valeurs extrêmes de l'année 2002 ; société CGBI 410
Tableau 144 : Valeurs extrêmes de l'année 2003 410
Tableau 145 : Valeurs extrêmes de l'année 2004 411
Tableau 146 : Ajustement de la distribution 1992 par une loi Logistique 412
Tableau 147 : Ajustement de la distribution 1993 par une loi Logistique 414
Tableau 148 : Ajustement de la distribution 1994 par une loi Logistique 415
Tableau 149 : Ajustement de la distribution 1995 par une loi Logistique 416
Tableau 150 : Ajustement de la distribution en 1996....417
Tableau 151 : Ajustement de la distribution en 1997....418
Tableau 152 : Ajustement de la distribution 1998 par une loi Logistique 419
Tableau 153 : Ajustement de la distribution 1999 par une loi Logistique 420
Tableau 154 : Ajustement de la distribution 2000 par une loi Logistique 421
Tableau 155 : Ajustement de la distribution 2001 par une loi Logistique 422
Tableau 156 : Ajustement de la distribution 2002 par une loi Logistique 423
Tableau 157 : Ajustement de la distribution 2003 par une loi Logistique 424
Tableau 158 : Ajustement de la distribution 2004 par une loi Logistique 425
Tableau 159 : Ajustement de la demi-distribution gauche 1992 par une loi Pareto2 426
Tableau 160 : Distribution des irrégularités en 1992 ; ajustement Pareto2....427

Tableau 161 : Ajustement de la demi-distribution gauche 1993 par une loi Pareto2 428
Tableau 162 : Ajustement de la demi-distribution gauche 1994 par une loi Pareto2 429
Tableau 163 : Ajustement de la demi-distribution gauche 1995 par une loi Pareto2 430
Tableau 164 : Ajustement de la demi-distribution gauche 1996 par une loi Pareto2 431
Tableau 165 : Ajustement de la demi-distribution gauche 1997 par une loi Pareto2 432
Tableau 166 : Ajustement de la demi-distribution gauche 1998 par une loi Pareto2 433
Tableau 167 : Ajustement de la demi-distribution gauche 1999 par une loi Pareto2 434
Tableau 168 : Ajustement de la demi-distribution gauche 2000 par une loi Pareto2 435
Tableau 169 : Ajustement de la demi-distribution gauche 2001 par une loi Pareto2 436
Tableau 170 : Ajustement de la demi-distribution gauche 2002 par une loi Pareto2 437
Tableau 171 : Ajustement de la demi-distribution gauche 2003 par une loi Pareto2 438
Tableau 172 : Ajustement de la demi-distribution gauche 2004 par une loi Pareto2 439
Tableau 173 : Irrégularité gauche (en 1992) mesurée en élargissant l'intervalle d'étude440
Tableau 174 : Irrégularité droite (en 1992) mesurée en élargissant l'intervalle d'étude ..441
Tableau 175 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1993 442
Tableau 176 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1994 443
Tableau 177 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1995 444
Tableau 178 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1996 445
Tableau 179 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1997 446
Tableau 180 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1998 447
Tableau 181 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 1999 448
Tableau 182 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2000 449
Tableau 183 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2001 450
Tableau 184 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2002 451
Tableau 185 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2003 452
Tableau 186 : Irrégularités droite et gauche mesurées en élargissant l'intervalle d'étude en 2004 454
Tableau 187 : Calcul du montant moyen par moyenne pondérée en 2001, 2002 et 2004 455

TABLE DES MATIÈRES DÉTAILLÉE

Remerciements 3
Plan simplifié de la thèse 4

INTRODUCTION GÉNÉRALE 6

PARTIE 1 : DÉFINITION DU CHAMP DE RECHERCHE 9

Chapitre I - Réflexion sur la valeur des informations comptables 10

A. L'origine de la question de recherche 10
1. Un constat initial… 10
2. …qui suscite des questions 11
B. Le cadre général dans lequel s'insère la recherche 12
1. La qualité des informations comptables 13
a. Résultat et qualité 13
b. Mesure de la qualité du résultat comptable 15
b.1. L'approche par les flux de trésorerie 16
b.2. L'approche par le marché 16
b.3. Autres approches 17
2. La fiabilité du résultat 18
a. Les manipulations comptables 19
b. La gestion du résultat 20
c. La mise en évidence et la mesure de la gestion du résultat 21
c.1. Méthodologies statistiques et méthodologies par étude de cas 22
c.2. Trois méthodologies statistiques 23

Chapitre II - Littérature sur les seuils 28

A. Les travaux observant les seuils 28
1. Une démarche empirique 29
2. Les différents seuils constatés 30
a. Les travaux précurseurs 30
b. Les travaux ultérieurs confirment unanimement l'existence des seuils 31
B. Les travaux sur les causes des seuils 32
1. L'irrégularité est la manifestation d'une manipulation 32
a. Le seuil du résultat nul : un postulat de manipulation 32
b. La manipulation est-elle marginale ? 33
c. Des explications alternatives 36
c.1. Les actifs financiers 36
c.2. La théorie de l'effort 37
c.3. Les contraintes réglementaires 38
c.4. La prudence des règles comptables 38
c.5. L'influence de la variable de mise à l'échelle 39
d. Le postulat est étendu aux autres seuils (prévisions et variations) 39

2. Les études des causes des irrégularités 40
a. Le cadre théorique de la recherche positive sur la gestion du résultat 41
Les coûts d'agence 41
Les coûts politiques 43
b. Les facteurs spécifiques comportementaux 44
c. Les coûts de transaction 46
d. L'arbitrage coût-avantage 48
e. L'influence de l'environnement 48
C. Les travaux étudiant les moyens et les conséquences 49
1. Les études des moyens mis en œuvre par les entreprises évitant les seuils 49
a. Seuils et flux de trésorerie 49
b. Seuils et accruals 50
c. Moyens spécifiques mis en œuvre 52
d. Utilisation des arrondis 52
e. Moyens mis en œuvre pour limiter les discontinuités 52
2. Les travaux étudiant les conséquences des seuils 53
a. Les entreprises qui évitent les petites pertes 53
b. Les entreprises qui évitent les erreurs de prévisions 54
D. Les travaux mesurant et comparant les seuils 55
1. Les hiérarchisations 56
2. Les études longitudinales 56
3. Les études internationales 57
E. Synthèse de la revue de littérature et définition des axes de recherche 60
1. Un cadre théorique incomplet 60
2. Une remise en cause du postulat de manipulation 61
3. Des limites méthodologiques 61
a. La largeur des intervalles d'observation 61
b. Le dénominateur utilisé pour « réduire » la variable résultat 62
c. La composition des échantillons 63
4. Définition des axes de recherche 64

Chapitre III - Cadre théorique explicatif de la gestion des seuils 66

A. Un éclairage terminologique conduisant à distinguer deux « effets de seuils » 67
1. Discontinuité, irrégularité et seuil 67
2. La notion d'effet de seuil 70
3. De l'observation des discontinuités à la gestion du résultat 72
B. Proposition d'un cadre d'analyse 75
1. Les motivations économiques de l'évitement des seuils 75
2. Les motivations opportunistes 77
3. Le double impact des facteurs psychologiques 78
4. Le rôle des anticipations 79
5. Synthèse et apports du cadre théorique 81
C. Conclusion de la partie 1 : problématique de la thèse 82

PARTIE 2 : IMPACT DES CHOIX MÉTHODOLOGIQUES 84

Chapitre IV – État des lieux et présentation de l'étude méthodologique 85

A. Impact de la méthode de mesure de l'irrégularité 86
1. Présentation du problème 86
a. Observation graphique des irrégularités de distribution 86

b. Mesures d'irrégularité ... 91
b.1- Mesures linéaires de l'amplitude ... 92
b.2- Mesure tenant compte de la non linéarité de la courbe ... 98
b.3- Mesures tenant compte des transferts d'observations ... 101
c. Les indicateurs d'asymétrie ... 103
c.1. L'indicateur d'asymétrie « d/g » (ou « r/l ») ... 104
c.2. L'indicateur « (r-l)/(r+l) » ... 106
c.3. Synthèse sur les indicateurs d'asymétrie ... 108
d. Synthèse des limites des mesures non paramétriques ... 109
2. Démarche de l'étude méthodologique ... 110
a. Mise en évidence de l'influence de la méthode retenue ... 110
b. Hiérarchisation des méthodes non paramétriques de mesure ... 111
b.1. La volatilité des mesures ... 111
b.2. La qualité de la régression ... 111
b.3. L'analyse de classification hiérarchique ... 111
b.4. La comparaison des tendances ... 112
b.5. La pertinence de l'indicateur en fonction de l'emplacement de l'irrégularité dans la distribution ... 112
c. Analyse de la pertinence des indicateurs d'asymétrie ... 114
d. Résultats attendus de l'étude empirique sur le choix de la méthode de mesure ... 114
B. Impact de la largeur des intervalles d'observation ... 115
1. Présentation du problème ... 115
2. Démarche de l'étude méthodologique ... 116
a. Démarche ... 116
b. Résultats attendus de l'étude empirique sur l'impact de la largeur des intervalles ... 116
C. Impact de la variable de mise à l'échelle ... 117
1. Présentation du problème ... 117
a. La valeur de marché ... 117
b. L'actif total ... 119
c. Le chiffre d'affaires ... 121
2. Démarche de l'étude méthodologique ... 122
a. Démarche ... 122
b. Résultats attendus de l'étude empirique sur l'impact du dénominateur ... 123

Chapitre V – Démarche générale de l'étude empirique et analyse de la base de données ... 124

A. Analyse préliminaire de la base initiale ... 124
1. Constitution de la base ... 124
a. Les entreprises inactives ... 124
b. Les entreprises de banque et assurance ... 126
c. Choix des Variables ... 126
d. Mise à l'échelle ... 127
e. Données internationales ... 127
f. Données trimestrielles ... 127
2. Synthèse de la base initiale ... 128
B. Correction de la base ... 129
1. Résultat strictement égal à zéro ... 129
2. Actif total strictement égal a zéro ... 129
3. Données récentes ... 129
4. Données manquantes ... 129
5. Valeurs extrêmes ... 130
C. Analyse descriptive de la base corrigée ... 131
1. Effectifs de l'échantillon ... 131

2. Évolution de la moyenne, la médiane et l'écart type 131
3. Composition de la base..... 134

Chapitre VI – Mise en œuvre de l'étude méthodologique 135

A. Impact de la méthode de mesure de l'irrégularité.....135
1. Étude des mesures fondées sur une estimation..... 135
a. Intervalle d'étude 136
a.1. Choix de la largeur de l'intervalle 136
a.2. Effectif des intervalles étudiés..... 138
b. Détail des calculs..... 139
c. Résultats ; comparaison des différentes mesures 149
c.1. Mesures de l'irrégularité gauche 150
c.1. Mesures de l'irrégularité droite 152
d. Hiérarchisation des mesures 153
d.1. La volatilité des mesures..... 154
d.2. La qualité de la régression 155
d.3. L'analyse de classification hiérarchique..... 156
d.4. La comparaison des tendances 158
d.5. Le choix théorique du meilleur outil..... 161
e. Conclusion et apports sur les mesures d'irrégularités 165
2. Étude des indicateurs d'asymétrie 167
a. Comparaison de l'évolution de l'irrégularité à droite et de l'irrégularité à gauche du seuil résultat nul 167
b. Mesure des indicateurs d'asymétrie et comparaison..... 170
c. Une relecture des travaux antérieurs..... 175
d. Conclusion et apports sur les indicateurs d'asymétrie..... 175
B. Impact de la largeur des intervalles, et nouvel indicateur d'irrégularité.....177
1. Choix d'une année d'étude et observation graphique 177
a. Choix de l'année 1997 177
b. Observations graphiques de la distribution en fonction de la largeur des intervalles 178
c. Limites des largeurs des intervalles..... 182
c.1. Contrainte pour le calcul par moyenne..... 182
c.2. Contrainte pour le calcul par symétrie 184
2. Évolution de la largeur des intervalles et irrégularité mesurée par moyenne..... 186
3. Évolution de la largeur des intervalles et irrégularité mesurée par symétrie 189
4. Comparaison des irrégularités mesurées par symétrie et moyenne 190
5. Proposition d'un nouvel indicateur de l'irrégularité 193
6. Conclusions et apports de l'étude de l'influence de la largeur des intervalles sur la mesure des irrégularités..... 196
C. Impact du choix de la variable de mise à l'échelle.....198
1. Distribution des résultats non réduits..... 198
Conclusion et apports sur la distribution des résultats non réduits..... 202
2. Comparaison des variables de mise à l'échelle..... 203
a. Les variables de mise à l'échelle 203
a.1. L'actif total 203
a.2. Le chiffre d'affaires 210
a.3. La valeur de marché..... 211
b. Analyse réduite aux observations communes 212
b.1. Étude des corrélations..... 213
b.2. Étude des corrélations sur les seules entreprises proches du seuil 214
c. Analyse étendue à toutes les observations disponibles 216
c.1. Comparaison des trois distributions et notion de petite performance 216
c.2. Comparaison des entreprises identifiées selon les trois variables..... 220
c.3. Mesure et comparaison des irrégularités 225
3. Conclusion et apports sur l'impact du choix du dénominateur 232

D. Conclusion de la partie 2 : Limites des mesures non paramétriques.....234

PARTIE 3 : PROPOSITION D'UN INSTRUMENT DE MESURE ET APPLICATION AUX ENTREPRISES FRANÇAISES COTÉES 237

Chapitre VII - Proposition d'un nouvel instrument de mesure des irregularités 238

A. Pourquoi faut-il adopter une méthode paramétrique ?.....238

B. Existe-t-il une loi de distribution des résultats ?.....241

1. Les irrégularités sous hypothèse de Normalité..... 241

a. Distribution du résultat non réduit..... 242

b. Distribution du résultat mis à l'échelle 244

b.1. Dissymétrie..... 245

b.2. Concentration 247

2. Théorème central limite..... 248

3. Synthèse et apports : La loi Normale n'est pas la loi de distribution 249

C. À la recherche d'une loi de distribution 250

1. Fondements de la démarche d'ajustement empirique de la loi 250

a. L'utilisation d'une caractéristique comptable 251

b. Le traitement mathématique adéquat..... 252

2. Mise en œuvre de la démarche 253

a. Utilisation de données pluriannuelles..... 253

b. Traitement des données supposées irrégulières (« zone inquiétante ») 254

3. Caractéristiques de la loi attendue..... 256

a. Caractéristiques attendues 256

b. Lois testées..... 256

b.1. Ajustement global 256

b.2. Ajustement partiel..... 258

c. Outils statistiques..... 259

c.1. Logiciel utilisé 259

c.2. Critères de qualité des ajustements 259

4. Finalité de la démarche..... 261

Chapitre VIII - Mise en œuvre : ajustement d'une loi de distribution des résultats des entreprises françaises..... 262

A. Valeurs extrêmes 262

1. Caractéristiques de la distribution..... 263

2. Règles de détermination des valeurs extrêmes..... 263

3. Analyse systématique de chaque année 264

B. Réduction.....266

C. Centrage.....267

D. Ajustement.....270

1. Paramètres de la distribution étudiée 270

2. Ajustement de la distribution globale (contrainte de symétrie) 272

a. Démarche 272

b. Synthèse de la démarche d'ajustement de la fonction globale..... 276

3. Ajustement de la demi-distribution (pas de contrainte de symétrie)..... 277

a. Démarche 277

a.1. Division en deux courbes 277

a.2. Jonction des deux courbes..... 278

a.3. La nature des lois testées ... 278
b. Ajustement des données positives (partie droite de la courbe) ... 279
c. Ajustement des données négatives (partie gauche de la courbe) ... 280
4. Extension de la méthode à d'autres populations : une étude internationale ... 282
a. Composition des échantillons internationaux ... 282
a.1. Analyse des données ... 283
a.2. Centrage des données ... 284
a.3. Description des résultats centrés réduits pour la Grande Bretagne ... 288
a.4. Description des résultats centrés réduits pour l'Allemagne ... 289
b. Ajustement d'une loi de distribution globale ... 290
b.1. Ajustement d'une loi de distribution des entreprises britanniques ... 290
b.2. Ajustement d'une loi de distribution des entreprises allemandes ... 292
b.3. Comparaison avec la distribution française ... 294
c. Ajustement d'une loi de distribution partielle (demi-distribution) ... 294
c.1. Ajustement d'une loi de distribution partielle des entreprises britanniques ... 294
c.2. Ajustement d'une loi de distribution partielle des entreprises allemandes ... 297
c.3. Comparaison avec la distribution française ... 300
5. Conclusion et apports de la démarche d'ajustement ... 301

Chapitre IX - Application dans le cadre d'une étude longitudinale ... 303

A. Étude longitudinale de l'évolution des seuils en France ... 303
1. Mesures des irrégularités ... 303
a. Mesures par la loi Logistique (hypothèse de symétrie) ... 304
a.1. La démarche générale ... 304
a.2. Les mesures année par année ... 304
a.3. La synthèse des résultats ... 306
b. Mesures par la loi Pareto2 (hypothèse de non symétrie) ... 308
b.1. La démarche générale ... 308
b.2. Les mesures année par année ... 308
La synthèse des résultats ... 310
2. Comparaison des mesures d'irrégularité ... 312
a. Comparaison des deux méthodes paramétriques ... 312
b. Comparaison des méthodes paramétriques et non paramétriques ... 313
B. Deux nouveaux indicateurs d'irrégularité ... 316
1. Démarche de l'étude ... 316
a. « Largeur » et « Hauteur » de l'irrégularité ... 316
b. Intérêt de ces deux nouveaux indicateurs ... 317
c. Détermination du seuil de significativité ... 319
c.1. Démarche générale ... 319
c.2. Test statistique de significativité ... 320
c.3. Détermination empirique d'un seuil de significativité ... 323
2. Mesures année par année ... 325
a. Largeur maximale année par année ... 326
a.1. La démarche générale ... 326
a.2. Largeur maximale en 1992 ... 326
a.3. Synthèse des mesures de largeurs max ... 328
b. Nombre total d'entreprises manipulatrices ... 329
c. Montant maximal manipulé ... 332
d. Montant moyen manipulé ... 335
d.1. détermination approximative du montant moyen manipulé ... 335
d.2. détermination du montant moyen manipulé par moyenne pondérée ... 337
d.3. Synthèse de l'étude des montant manipulés ... 339
e. Modification du seuil de significativité ... 341
3. Synthèse de l'étude longitudinale ... 343
C. Conclusion de la partie 3 : Les seuils comptables en France ... 347

CONCLUSION GÉNÉRALE 351

De l'importance des seuils comptables 351
Apport sémantique 353
Apport méthodologique 354
Apport empirique 355
En conclusion… 356

ANNEXES 357

Annexe 1 : Correction de la base 358
1. Résultat strictement égal à zéro 358
2. Actif total strictement égal à zéro 359
3. Données récentes 359
4. Données manquantes 360
5. Valeurs extrêmes 367
Annexe 2 : Équations de régressions 373
Annexe 3 : Emplacement de l'irrégularité dans la distribution annuelle et choix de la méthode de mesure appropriée 376
Annexe 4 : Observation des intervalles encadrant le seuil, et mesures des irrégularités par moyenne en faisant varier la largeur des intervalles 380
Annexe 5 : Mesures des irrégularités par symétrie en faisant varier la largeur des intervalles 382
Annexe 6 : Distributions annuelles de la variable résultat non mise à l'échelle 384
Annexe 7 : Détail de l'examen de la variable CA dans la base 389
Annexe 8 : Largeurs théoriques en fonction du dénominateur 392
Annexe 9 : Irrégularités mesurées en mettant le résultat à l'échelle par le chiffre d'affaires 393
Annexe 10 : Irrégularités mesurées en mettant le résultat à l'échelle par la valeur de marché 398
Annexe 11 : Second examen des valeurs extrêmes au cas par cas 403
Annexe 12 : Mesures année par année avec une loi Logistique (hypothèse de symétrie) . 412
Annexe 13 : Mesures année par année avec une loi Pareto 2 (hypothèse de non symétrie) 426
Annexe 14 : Mesure de la largeur maximale année par année 440
Annexe 15 : Calcul du montant moyen annuel manipulé par les entreprises qui évitent une perte ; calcul par moyenne pondérée 455

BIBLIOGRAPHIE ET INDEX DES FIGURES ET TABLEAUX 457

Bibliographie 457
Index des figures 464
Index des tableaux 469

TABLE DES MATIÈRES DÉTAILLÉE 475

Résumé

Titre : Gestion du résultat et seuils comptables : Impact des choix méthodologiques et proposition d'un instrument de mesure des irrégularités.

La thèse étudie l'influence des seuils comptables dans les pratiques de gestion du résultat. Ses contributions sont (1) sémantique, (2) méthodologique et (3) empirique.

(1) La thèse clarifie la terminologie et dresse une typologie des seuils. Elle propose un cadre d'analyse qui complète le modèle politico-contractuel par des explications psychologiques et met en relief le rôle clef des anticipations des dirigeants.

(2) Les mesures non paramétriques habituellement utilisées pour étudier les irrégularités de distribution des résultats publiés sont examinées. Trois limites sont identifiées : le choix d'un instrument de mesure, de la largeur d'intervalle et de la variable de mise à l'échelle. Leur impact est évalué par une étude empirique des entreprises cotées françaises. Pour dépasser ces limites, la thèse propose une démarche paramétrique qui permet de mesurer le nombre total d'entreprises manipulatrices et les montants manipulés.

(3) La démarche paramétrique est mise en œuvre pour étudier les résultats des entreprises françaises cotées de 1992 à 2004. Elle révèle que 5 à 10 % des entreprises manipulent chaque année leur résultat pour éviter de publier une perte. Les montants annuels manipulés sont évalués à 8,5 milliards d'euros et diminuent au cours de la période.

En remettant en cause le postulat de manipulation marginale (ou évitement des petites pertes), la thèse souligne l'importance de la recherche sur les seuils, et ce résultat, observé en France, dépasse le cadre national.

Mots clefs : gestion du résultat, seuils comptables, théorie politico-contractuelle, distribution des résultats, modèle paramétrique, irrégularité de distribution, manipulation marginale

Abstract

Title: Earnings management and accounting thresholds: Influence of methodological choices and proposition of a measuring instrument for irregularities.

This research analyzes the influence of accounting thresholds in earnings management practices. Its contributions are threefold: (1) semantic, (2) methodological and (3) empirical.

(1) The research clarifies the terminology and suggests a typology of thresholds. It proposes an additional framework for positive accounting theory by considering psychological explanations and highlighting the importance of managers' anticipations.

(2) The study investigates nonparametric models mobilized in previous studies on earnings distributions. It identifies three limits: the estimating for irregularities, the interval width, and the scaling variable. The consequences of these limits are assessed by an empirical study of French listed companies. Findings call for a parametric model to overcome these deficiencies. This model allows measuring the total number of companies which manage their earnings, and the amounts they manage.

(3) This parametric model is implemented to analyze the earnings distributions of French listed companies from 1992 to 2004. Results show that each year, 5 to 10% of the companies manage their earnings to avoid disclosing a loss. Annual managed amounts are estimated at about 8.5 billion euros and decrease during the period.

By questioning the postulate of small losses avoidance, this research stresses the importance of thresholds research, and calls for further international studies.

Keywords: Earnings Management, Accounting Thresholds, Positive Accounting Theory, Earnings Distribution, Parametric Model, Distribution Irregularity, Small Losses Avoidance