



**HAL**  
open science

## Une Evaluation des Procédures de Backtesting

Christophe Hurlin, Sessi Tokpavi

► **To cite this version:**

Christophe Hurlin, Sessi Tokpavi. Une Evaluation des Procédures de Backtesting: Tout va pour le Mieux dans le Meilleur des Mondes. 2007. halshs-00159846

**HAL Id: halshs-00159846**

**<https://shs.hal.science/halshs-00159846>**

Preprint submitted on 4 Jul 2007

**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.

# Une Evaluation des Procédures de Backtesting

"Tout va pour le Mieux dans le Meilleur des Mondes"

Christophe Hurlin <sup>\*</sup> and Sessi Tokpavi <sup>†</sup>

Juin 2007

## Abstract

Dans cet article, nous proposons une démarche originale visant à évaluer la capacité des tests usuels de backtesting à discriminer différentes prévisions de Value at Risk (VaR) ne fournissant pas la même évaluation *ex-ante* du risque. Nos résultats montrent que, pour un même actif, ces tests conduisent très souvent à ne pas rejeter la validité, au sens de la couverture conditionnelle, de la plupart des six prévisions de VaR étudiées, même si ces dernières sont sensiblement différentes. Autrement dit, toute prévision de VaR a de fortes chances d'être validée par ce type de procédure.

## Abstract

This paper proposes an evaluation of backtests that examine the accuracy of Value-at-Risk (VaR) forecasts. It is well known that VaR backtesting procedures outlined by the Basel Committee for Banking Supervision have limited power to control the probability of accepting an incorrect VaR forecast. In this study, we propose an original approach based on the replication of these tests on six different VaR forecasts (parametric or non parametric) for a given asset. We show that backtests generally lead to not reject the accuracy of all (or most of) these different forecasts. In other words, most of VaR forecasts are likely to be considered as valid.

- *Mots-Clés* : Value-at-Risk; Backtesting
- *J.E.L Classification* : C22, C52, G28

---

<sup>\*</sup>LEO, Université d'Orléans. Rue de Blois. BP 6739. 45067 Orléans Cedex 2. France. e-mail address: christophe.hurlin@univ-orleans.fr.

<sup>†</sup>LEO, Université d'Orléans. Rue de Blois. BP 6739. 45067 Orléans Cedex 2. France. e-mail address: sessi.tokpavi@univ-orleans.fr.

# 1 Introduction

"*Tout va pour le mieux dans le meilleur des mondes*" est la phrase qui clôt Candide après que le protagoniste et son maître à penser Pangloss aient pourtant connu toutes les vicissitudes de la nature humaine. Telle pourrait être aussi la conclusion de notre évaluation systématique des procédures de backtesting de la Value-at-Risk (VaR). Notre étude montre en effet que pour un même actif, plusieurs prévisions différentes de la VaR et donc du risque de portefeuille, ont de très fortes chances d'être considérées conjointement comme valides à l'issue des procédures usuelles de backtesting, dans une sorte d'optimisme que ne renierait pas Leibniz.

Plus précisément, nous proposons dans cet article une démarche originale visant à évaluer la capacité des tests standards de backtesting à discriminer différentes prévisions de VaR ne fournissant pas la même évaluation *ex-ante* du risque de portefeuille. Une prévision de VaR est dite valide si la séquence des violations<sup>1</sup> associées satisfait les hypothèses de couverture non conditionnelle et d'indépendance (Christoffersen, 1998). Plusieurs approches sont aujourd'hui proposées dans la littérature (voir Campbell, 2007, pour un survey) pour tester ces deux hypothèses de façon séparée ou de façon jointe (hypothèse de couverture conditionnelle). La démarche habituelle d'évaluation de ces procédures de backtesting consiste à mesurer leur taille et leur puissance empiriques sur des échantillons simulés<sup>2</sup> de différentes dimensions (Berkowitz, 2001; Christoffersen et Pelletier,

---

<sup>1</sup>On définit la violation comme une variable dichotomique prenant la valeur un lorsque le rendement de l'actif observé *ex-post* inférieur à la VaR prédite *ex-ante* et zéro dans le cas contraire.

<sup>2</sup>Dans le cadre par exemple d'une étude de la puissance, ce type d'exercice consiste à se donner un processus générateur de données pour la distribution conditionnelle des rentabilités (un

2004; Berkowitz, Christoffersen et Pelletier, 2005; etc.).

La méthode d'évaluation des procédures de backtesting que nous avons retenue ici est sensiblement différente. Nous proposons de vérifier si, lorsque l'on dispose de plusieurs prévisions différentes de la VaR pour un même portefeuille, ces tests permettent de distinguer un sous ensemble de prévisions valides (respectivement non valides). En effet, il existe aujourd'hui une très grande variété de méthodes paramétriques ou non paramétriques de calcul de la VaR (voir Downd, 2005 pour une synthèse). Or, il est reconnu que ces méthodes fournissent généralement pour un même portefeuille des mesures significativement différentes de la VaR et donc du risque (Beder, 1995). Dès lors, la question se pose de savoir si les procédures de backtesting usuelles permettent de discriminer ces différentes prévisions. Il ne s'agit pas ici d'évaluer la capacité de ces tests à rejeter l'hypothèse de couverture conditionnelle, puisque sur données réelles la «vraie» mesure de VaR est a priori inconnue. Il s'agit uniquement de vérifier à partir d'un grand nombre d'expériences, si parmi  $N$  prévisions différentes de VaR issues de  $N$  méthodes de calcul alternatives, les tests permettent effectivement de rejeter la validité d'au moins quelques unes de ces prévisions.

Pour cette étude, nous considérons trois tests qui sont parmi les plus utilisés et les plus souvent cités dans la littérature consacrée au backtesting : les tests de ratio de vraisemblance de Christoffersen (1998), les tests de quantiles dynamiques (*Dynamic Quantil*,  $DQ$ ) d'Engle et Manganelli (2004) et les tests de durée de

---

processus GARCH par exemple) et à retenir une méthode de calcul de la VaR (méthode hybride, méthode historique, etc.) qui ne permet pas dans ce cas de modéliser correctement les fractiles de la distribution conditionnelle. A partir de simulations de Monte Carlo du processus des rentabilités, en répliquant la méthode de calcul de VaR, on cherche alors à évaluer la fréquence de rejets de l'hypothèse de couverture conditionnelle pour différentes tailles d'échantillons.

Christoffersen et Pelletier (2004). Ces trois tests et leurs déclinaisons seront systématiquement appliqués pour discriminer parmi six mesures de VaR incluant trois mesures paramétriques (méthode Delta Normale, modèle GARCH et modèle RiskMetrics), deux mesures non paramétriques (méthode hybride et méthode de simulation historique) et une mesure semi-paramétrique (modèle CAViaR). Afin d'évaluer la robustesse de nos résultats ces expériences sont réalisées sur les rendements associés à trois types d'actifs : l'indice Nasdaq, le titre General Motors et le taux de change Mark-Dollar US.

Le plan du reste de l'article est le suivant. Dans une première section, nous définissons les principales notions de validité de la VaR (couverture non conditionnelle, indépendance et couverture conditionnelle) et nous présentons les principaux tests de backtesting associés. Dans une seconde section, nous décrivons en détail notre démarche d'évaluation en insistant en particulier sur les méthodes de calcul de la VaR. Enfin, dans une troisième section nous présentons les résultats.

## **2 Les Principaux Tests de Validation de la VaR**

Traditionnellement, la qualité de la prévision d'une grandeur économique est évaluée en comparant sa réalisation *ex-post* à la valeur prédite *ex-ante*. Le calcul des erreurs de prévision permet alors de construire des critères d'évaluation des qualités prédictives des différents modèles, comme par exemple le critère de la *Root Mean Square Forecast Error (RMSFE)* ou celui de la *Mean Absolute Error (MAE)*. L'observation de ces erreurs rend en outre possible la construction d'un test de l'hypothèse nulle selon laquelle deux modèles concurrents présentent les mêmes qualités prédictives (mesurées en termes de *RMSFE* ou de tout autre

critère) dans une approche à la Diebold et Mariano (1995). Naturellement, ces procédures de comparaison de modèles ne sont applicables que si la réalisation *ex-post* de la variable d'intérêt est observable. Dans le cas contraire, l'exercice d'évaluation nécessite alors d'utiliser une variable proxy de la variable d'intérêt latente. Un exemple assez connu est celui de l'évaluation des modèles de volatilité, où la volatilité journalière *ex-post* peut être approximée de façon satisfaisante par la volatilité réalisée, définie comme la somme des carrés des rentabilités intra-day (Anderson et al., 2003).

Toutefois, dans le cadre de la problématique VaR, la construction d'une telle variable proxy s'avère relativement délicate. C'est pourquoi les critères d'évaluation des prévisions de la VaR sont généralement fondés sur des tests statistiques (et non de simples critères) des deux principales hypothèses que le modèle de mesure du risque doit satisfaire : l'hypothèse de couverture non conditionnelle et l'hypothèse d'indépendance des violations.

On note  $VaR_{t|t-1}(\alpha)$  la prévision de VaR par unité monétaire pour la période  $t$  établie conditionnellement à l'ensemble d'information  $\Omega_{t-1}$  disponible à la date  $t-1$  et pour un taux de couverture de  $\alpha\%$ . Par définition, cette quantité correspond au fractile d'ordre  $\alpha$  de la distribution conditionnelle des rendements  $r_t$  de l'actif ou du portefeuille d'actifs considéré :

$$\Pr [r_t < VaR_{t|t-1}(\alpha)] = \alpha \quad (1)$$

Une violation de la VaR (ou *hit*) apparaît lorsque la rentabilité observée *ex-post*,  $r_t$ , est inférieure à la valeur prévue *ex-ante* de la VaR. On note  $I_t(\alpha)$  la variable indicatrice associée à une éventuelle violation de la VaR à la date  $t$  pour

un taux de couverture de  $\alpha\%$  :

$$I_t(\alpha) = \begin{cases} 1 & \text{si } r_t < VaR_{t|t-1}(\alpha) \\ 0 & \text{sinon} \end{cases} \quad (2)$$

Christoffersen (1998) montre que le problème de la validité de la prévision de VaR peut se ramener au problème de savoir si la séquence des violations  $\{I_t(\alpha)\}_{t=1}^T$  satisfait ou non les deux hypothèses suivantes:

- L'hypothèse de couverture non conditionnelle : la probabilité que se réalise *ex-post* une perte en excès par rapport à la VaR anticipée *ex-ante* doit précisément être égale au taux de couverture  $\alpha$  :

$$E[I_t(\alpha)] = \Pr[I_t(\alpha) = 1] = \alpha \quad (3)$$

- L'hypothèse d'indépendance : les violations de la VaR pour un même taux de couverture à deux dates différentes doivent être indépendamment distribuées. Formellement, la variable  $I_t(\alpha)$  associée à la violation à la date  $t$  de la VaR pour un taux de couverture à  $\alpha\%$ , est indépendante de la variable  $I_{t-k}(\alpha)$  pour toute valeur de  $k$  différente de 0. Autrement dit, l'observation des violations passées de la VaR n'apporte aucune information sur les violations contemporaines et futures. Cette propriété est en outre valable pour n'importe quelle variable appartenant à l'ensemble d'information disponibles à la date  $t - 1$ .

La première hypothèse est tout à fait intuitive : si la prévision de VaR est valide, la fréquence des violations observées sur une période  $T$  où cette VaR est retenue comme mesure du risque extrême, *i.e.*  $(1/T) \sum_{t=1}^T I_t(\alpha)$ , ne doit pas être significativement différente du taux de couverture nominale  $\alpha$ . Dans le

cas contraire, si la fréquence des violations est significativement inférieure (resp. supérieure) au taux de couverture nominal  $\alpha$ , cela traduit une surestimation (resp. sous-évaluation) de la VaR et donc du risque. Les tests de validité de cette propriété de couverture non conditionnelle, initialement développés par Kupiec (1995), constituent aujourd'hui le coeur des principales procédures d'évaluation des modèles de VaR, préconisées notamment au niveau réglementaire.

Toutefois, si la couverture non conditionnelle permet de s'assurer que la proportion de violations pour une période donnée garantit la couverture nominale, elle ne donne aucune information sur l'indépendance temporelle des violations. Or, la propriété d'indépendance des violations est une propriété essentielle puisque toute mesure de risque doit s'ajuster sans retard à toute nouvelle information entraînant une évolution dans la dynamique de la rentabilité de l'actif. Une modélisation qui ne prend pas en compte cet aspect, risque d'engendrer des clusters de violations successives. Aussi, aucune forme de dépendance ne doit elle exister dans la séquence des violations et cela quels que soient les taux de couverture considérés.

Il convient de noter que ces deux propriétés de la VaR sont indépendantes l'une de l'autre. Dès lors, si une prévision de VaR ne satisfait pas à l'une ou l'autre de ces deux hypothèses, elle doit être considérée comme non valide. A l'inverse, on qualifie de couverture conditionnelle la situation qui prévaut lorsque la prévision de VaR satisfait les deux hypothèses. Dans le cadre des procédures de backtesting, de nombreux tests ont été proposés dans la littérature pour tester l'une ou l'autre des propriétés de couverture non conditionnelle et d'indépendance, ou directement la propriété de couverture conditionnelle. Dans ce contexte, la



difficulté majeure consiste à spécifier la forme de la dépendance des processus  $I_t(\alpha)$  sous l'hypothèse alternative. Aux différentes spécifications proposées sont associés différents tests parmi lesquels nous ne retiendrons ici que trois approches qui constituent aujourd'hui les références de la littérature : le test LR de Christoffersen (1998), le test DQ d'Engle et Manganelli (2004) et le test de modèle de durée de Christoffersen et Pelletier (2004).

## 2.1 Test LR de Christoffersen (1998)

Christoffersen (1998) propose un test dans lequel on suppose que, sous l'hypothèse alternative de non efficacité de la VaR, le processus des violations  $I_t(\alpha)$  est modélisé par une chaîne de Markov admettant pour matrice des probabilités de transition la matrice suivante :

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} \quad (4)$$

où  $\pi_{ij} = \Pr[I_t(\alpha) = j \mid I_{t-1}(\alpha) = i]$ . De façon générale, cette chaîne de Markov permet de modéliser une éventuelle dépendance temporelle dans la séquence  $I_t(\alpha)$ . Ainsi, la probabilité d'observer une violation (resp. de ne pas en observer) pour la période courante peut être liée (ou non) à l'occurrence d'une violation (pour un même niveau de couverture  $\alpha$ ) à la période précédente. L'hypothèse nulle d'efficacité conditionnelle est alors définie par l'égalité :

$$H_0 : \Pi = \Pi_\alpha = \begin{pmatrix} 1 - \alpha & \alpha \\ 1 - \alpha & \alpha \end{pmatrix} \quad (5)$$

L'acceptation de l'hypothèse nulle entraîne le respect de l'hypothèse de couverture non conditionnelle. Quel que soit l'état du système en  $t - 1$ , la probabilité d'observer à la date  $t$  une violation est égale au taux de couverture conditionnelle,

*i.e.*  $\pi_t = \Pr [I_t(\alpha) = 1] = \alpha$ . De plus, la probabilité d'observer une violation à la date  $t$  est indépendante de l'état en  $t - 1$ . Une simple statistique de rapport de vraisemblance, notée  $LR_{CC}$ , permet alors de tester l'hypothèse nulle d'efficience conditionnelle. Sous  $H_0$ , on montre que :

$$LR_{CC} = -2 \left\{ \ln L [\Pi_\alpha, I_1(\alpha), \dots, I_T(\alpha)] - \ln L [\widehat{\Pi}, I_1(\alpha), \dots, I_T(\alpha)] \right\} \xrightarrow[T \rightarrow \infty]{L} \chi^2(2) \quad (2)$$

où  $\widehat{\Pi}$  désigne l'estimateur du maximum de vraisemblance de la matrice de transition sous l'hypothèse alternative et où  $\ln L [\Pi, I_1(\alpha), \dots, I_T(\alpha)]$  désigne la log-vraisemblance des violations  $I_t(\alpha)$  associées à une matrice de transition  $\Pi$ . Par définition :

$$L [\Pi, I_1(\alpha), \dots, I_T(\alpha)] = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}} \quad (6)$$

où  $n_{ij}$  désigne le nombre de fois où l'on observe  $I_t(\alpha) = j$  sachant  $I_{t-1}(\alpha) = i$ . Dans cette perspective, il est en outre possible de distinguer suivant que l'inefficience est due au non respect de l'hypothèse d'indépendance et / ou au non respect de l'hypothèse de couverture non conditionnelle. Ainsi, Christoffersen propose deux autres statistiques de rapport de vraisemblance, notée  $LR_{UC}$  et  $LR_{IND}$ , associées à chacune de ces deux sous hypothèses. Ainsi, sous l'hypothèse nulle d'indépendance, la matrice de transition est définie par :

$$H_0 : \Pi = \Pi_\pi = \begin{pmatrix} 1 - \pi & \pi \\ 1 - \pi & \pi \end{pmatrix} \quad (7)$$

où la probabilité  $\pi$  n'est pas nécessairement égale au taux de couverture non conditionnelle  $\alpha$ . Dès lors, la statistique  $LR_{IND}$  associée à la seule hypothèse nulle d'indépendance des violations est définie par :

$$LR_{IND} = -2 \left\{ \ln L [\widehat{\Pi}_\pi, I_1(\alpha), \dots, I_T(\alpha)] - \ln L [\widehat{\Pi}, I_1(\alpha), \dots, I_T(\alpha)] \right\} \xrightarrow[T \rightarrow \infty]{L} \chi^2(1)$$

où  $\widehat{\Pi}_\pi$  désigne l'estimateur du maximum de vraisemblance de la matrice de transition sous l'hypothèse d'indépendance.

Ces tests sont très faciles à mettre en oeuvre, mais il apparaissent toutefois assez réducteurs pour deux raisons essentielles. Tout d'abord, l'indépendance est testée contre une forme très particulière qui ne prend pas en compte notamment des dépendances d'ordre supérieur à un. De plus, l'utilisation d'une chaîne de Markov ne permet pas de mesurer le rôle d'autres variables que la seule séquence des violations passées  $I_t(\alpha)$  dans une possible dépendance des violations. Le test récemment proposé par Engle et Manganelli (2004) permet de lever ces deux inconvénients.

## 2.2 Test DQ de Engle et Manganelli (2004)

Engle et Manganelli (2004) proposent d'utiliser un modèle de régression linéaire liant les violations courantes aux violations passées afin de tester l'hypothèse d'efficience conditionnelle. Soit  $Hit(\alpha) = I_t(\alpha) - \alpha$ , le processus de violations centré sur  $\alpha$  associé à  $I_t(\alpha)$  :

$$Hit_t(\alpha) = \begin{cases} 1 - \alpha & \text{si } r_t < VaR_{t|t-1}(\alpha) \\ -\alpha & \text{sinon} \end{cases} \quad (8)$$

Considérons le modèle de régression linéaire suivant :

$$Hit_t(\alpha) = \delta + \sum_{k=1}^K \beta_k Hit_{t-k}(\alpha) + \sum_{k=1}^K \gamma_k g[Hit_{t-k}(\alpha), Hit_{t-k-1}(\alpha), \dots, z_{t-k}, z_{t-k-1}, \dots] + \varepsilon_t \quad (9)$$

où  $\varepsilon_t$  est un processus *i.i.d.* et où  $g(\cdot)$  désigne une fonction des violations passées et de variables  $z_{t-k}$  appartenant à l'ensemble d'information disponible  $\Omega_{t-1}$ . On peut ici considérer par exemple les rentabilités passées  $r_{t-k}$ , le carré des rentabilités

passées  $r_{t-k}^2$ , les valeurs passées de la VaR,  $VaR_{t-k|t-k-1}(\alpha)$  ou encore des données de volatilités implicites. Mais, quelle que soit la spécification retenue, le test de l'hypothèse nulle d'efficience conditionnelle se ramène à tester la nullité jointe des coefficients  $\beta_k$  et  $\gamma_k$  et de la constante  $\delta$  :

$$H_0 : \delta = \beta_k = \gamma_k = 0, \quad \forall k = 1, \dots, K \quad (10)$$

Les violations courantes de la VaR sont non corrélées aux violations passées dès lors que  $\beta_k = \gamma_k = 0$  (implication de l'hypothèse d'indépendance), tandis que l'hypothèse de couverture non conditionnelle est satisfaite dès lors que la constante  $\delta$  est nulle. En effet, sous l'hypothèse nulle,  $E[Hit_t(\alpha)] = E(\varepsilon_t) = 0$ , ce qui implique par définition que  $\Pr[I_t(\alpha) = 1] = E[I_t(\alpha)] = \alpha$ . Le test de nullité jointe de tous les coefficients correspond donc à un test d'efficience conditionnelle. Une statistique LR ou une statistique de Wald peuvent facilement être mises en oeuvre pour tester la nullité simultanée de ces coefficients. Ainsi, si l'on note  $\Psi = (\delta \ \beta_1 \dots \beta_K \ \gamma_1 \dots \gamma_K)'$  le vecteur des  $2K + 1$  paramètres de ce modèle et  $Z$  la matrice des variables explicatives du modèle (9), la statistique de Wald, notée  $DQ_{CC}$ , associée au test de l'hypothèse d'efficience conditionnelle<sup>3</sup> vérifie alors :

$$DQ_{CC} = \frac{\widehat{\Psi}' Z' Z \widehat{\Psi}}{\alpha(1-\alpha)} \xrightarrow[T \rightarrow \infty]{L} \chi^2(2K+1) \quad (11)$$

De la même façon que pour le test de Christoffersen, on peut bien évidemment décomposer ce test en ne testant par exemple que l'hypothèse d'indépendance des violations. On peut ainsi construire une statistique  $DQ_{IND}$  associée au test de l'hypothèse d'indépendance  $H_0 : \beta_k = \gamma_k = 0$  qui vérifie :

$$DQ_{IND} = \frac{\widehat{\Psi}' R' [R(Z'Z)^{-1} R']^{-1} R \widehat{\Psi}}{\alpha(1-\alpha)} \xrightarrow[T \rightarrow \infty]{L} \chi^2(2K) \quad (12)$$

---

<sup>3</sup>Sous l'hypothèse nulle, le résidu  $\varepsilon_t$  correspond au processus de violation  $Hit_t(\alpha)$  qui suit une loi de Bernouilli de paramètre  $\alpha$  de variance  $\alpha(1-\alpha)$ .

où  $R = [0 : I_{2K}]$  est une matrice de dimension  $(2K, 2K + 1)$  telle que  $R\Psi = \beta$ , avec  $\beta = (\beta_1 \dots \beta_K \ \gamma_1 \dots \gamma_K)'$ .

### 2.3 Test de durée de Christoffersen et Pelletier (2004)

Une approche alternative des tests de validation de la VaR repose sur la modélisation de la durée entre deux violations. C'est en particulier la démarche adoptée dans le cas du test de Christoffersen et Pelletier (2004). Ce test exploite l'idée que l'indépendance des violations de la VaR (c'est à dire l'absence de cluster des violations) implique que la durée entre deux violations n'a pas de mémoire et a une moyenne égale à  $1/\alpha$ . Ce faisant, Christoffersen et Pelletier évitent le problème de la spécification de la forme de la dépendance des violations, mais retrouvent un problème similaire qui consiste à postuler une durée entre deux violations en présence de dépendance. Formellement, si l'on note  $D_i = t_i - t_{i-1}$  la durée entre la  $i^{\text{ème}}$  violation et celle qui la précède, pour  $i = 1, \dots, T$ . Sous l'hypothèse nulle d'indépendance des violations, la durée  $D_i$  suit une loi exponentielle de paramètre  $\alpha$  et de densité :

$$f(d; \alpha) = \alpha \exp(-\alpha d) \quad (13)$$

Sous l'hypothèse alternative, Christoffersen et Pelletier supposent que la durée suit une loi de Weibull de paramètre d'échelle  $b$ , de paramètre de centrage  $a$  et de densité égale à :

$$g(d; b, a) = a^b b d^{b-1} \exp\left[-(ad)^b\right] \quad (14)$$

Cette hypothèse permet alors de proposer un test paramétrique simple de l'hypothèse d'indépendance puisque lorsque le paramètre d'échelle  $b$  est égal à 1, la loi de Weibull se ramène à une loi exponentielle. En outre, l'hypothèse

de couverture conditionnelle implique une contrainte sur la durée moyenne entre deux violations (égale à l'inverse du taux de couverture) et donc sur le paramètre de centrage  $a$ . Ainsi, le test de l'hypothèse d'efficience conditionnelle se ramène au test joint des hypothèses  $b = 1$  et  $a = \alpha$ . Une mise en oeuvre possible du test consiste à construire une statistique de ratio de vraisemblance définie par :

$$LR_{CC}^{duree} = 2 (ll_{nc} - ll_c) \xrightarrow{T \rightarrow \infty} \chi^2(2) \quad (15)$$

où  $ll_c$  et  $ll_{nc}$  désignent les log-vraisemblances respectives obtenues pour la loi exponentielle et la loi de Weibull.

### 3 Méthodes de Prédiction de la VaR

Comme nous l'avons dit précédemment, l'objectif de notre étude est de vérifier si, pour un actif donné, lorsque l'on dispose d'un grand nombre de méthodes alternatives de prédiction de la VaR, les tests standards de validation permettent de repérer un ensemble de prévisions valides du risque (au sens de la couverture conditionnelle). Afin d'évaluer précisément cette capacité à discriminer, nous proposons d'appliquer les tests de backtesting sur différents types d'actifs pour lesquels nous appliquerons à chacun six méthodes alternatives de prédiction de VaR.

Plus précisément, nous considérons six méthodes de prédiction relevant d'approches différentes afin de garantir l'obtention de mesures de risques divergentes. Ainsi, sont retenues trois approches paramétriques (méthode Delta Normale, modèle RiskMetrics et modèle GARCH), deux approches non paramétriques (simulation historique et méthode hybride) et une méthode semi-paramétrique (CAViaR), qui figurent pour parmi les plus utilisées dans la pratique.

Parmi les trois approches paramétriques retenues, la méthode Delta Normale et le modèle RiskMetrics sont fondées sur l'hypothèse de normalité de la distribution des rendements. Dans le cas de la méthode Delta Normale, cette hypothèse porte sur la distribution non conditionnelle des rendements. La VaR à  $\alpha\%$  calculée conditionnellement à l'information disponible à la période  $t - 1$  est alors définie par la relation :

$$VaR_{t|t-1}(\alpha) = \Phi^{-1}(\alpha) [var(\{r_j\}_{j=t-T_e}^{t-1})]^{1/2} + mean(\{r_j\}_{j=t-T_e}^{t-1}) \quad (16)$$

où  $\Phi(\cdot)$  désigne la fonction de répartition de loi normale centrée réduite et où la variance de la distribution non conditionnelle des rendements  $r$  est estimée à partir des  $T_e$  observations passées les plus récentes. Dans le cas de la méthodologie RiskMetrics, l'hypothèse de normalité porte cette fois sur la distribution conditionnelle des rendements. On considère un processus de variance conditionnelle hétéroscédastique qui s'apparente à la définition d'un IGARCH (Engle et Bollerslev, 1986) :

$$r_t = c + \varepsilon_t \quad (17)$$

$$\sigma_t^2 = (1 - \gamma) \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 \quad (18)$$

avec  $\varepsilon_t = z_t \sigma_t$  où  $z_t$  désigne un bruit blanc faible *N.i.d*  $(0, 1)$ . Les paramètres du modèle RiskMetrics sont généralement étalonnés et non estimés, avec en particulier une valeur de  $\gamma$  fixée à 0.94. La VaR à  $\alpha\%$  est calculée à partir de la variance conditionnelle  $\sigma_t^2$  pour la période  $t$  selon la formule suivante :

$$VaR_{t|t-1}(\alpha) = \Phi^{-1}(\alpha) \sigma_t + c \quad (19)$$

Toutefois, il est reconnu que cette hypothèse de normalité des résidus standardisés n'est pas satisfaisante puisqu'elle ne permet pas de prendre en compte

les caractéristiques leptokurtiques des distributions de rendements des actifs financiers<sup>4</sup>. Le recours à des distributions leptokurtiques constitue alors une solution. Dans cette perspective, nous considérons une troisième méthode paramétrique de calcul de la VaR fondée sur l'utilisation d'un modèle GARCH estimée par maximum de vraisemblance sous l'hypothèse d'une distribution de Student des résidus standardisés. Dans ce cas, la prévision de la variance conditionnelle est établie à partir du modèle suivant :

$$r_t = c + \varepsilon_t \quad (20)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (21)$$

avec  $\varepsilon_t = z_t \sigma_t$  où  $z_t$  désigne un bruit blanc faible distribué selon une loi de Student de paramètre  $v > 2$ . Les paramètres  $v, \alpha_0 > 0$  et  $(\alpha_1, \beta_1) \in \mathbb{R}^+$  avec  $\alpha_1 + \beta_1 < 1$  sont estimés de façon convergente par maximum de vraisemblance. La VaR à  $\alpha\%$  est alors calculée selon la relation :

$$VaR_{t|t-1}(\alpha) = F^{-1}(\alpha, v) \sigma_t + c \quad (22)$$

où  $F(z, v)$  désigne la fonction de répartition de la loi de Student à  $v$  degrés de liberté.

A l'inverse l'utilisation de mesures de VaR non paramétriques, comme la simulation historique (HS) et ses variantes, permet de s'affranchir du choix d'une quelconque distribution des rendements. Formellement la VaR HS est approximée par le quantile empirique d'ordre  $\alpha$  de la chronique des rentabilités passées

---

<sup>4</sup>Rappelons que dans cette étude, notre approche nécessite que l'on utilise à la fois des "bonnes" et des "mauvaises" mesures de risque. Dès lors, l'utilisation de mesures paramétriques fondées sur une hypothèse de normalité, sans doute non valide en pratique, nous garantit a priori d'obtenir de "mauvaises" mesures du risque qui devraient être identifiées par les tests de backtesting.



observées sur une fenêtre de taille  $Te$ , soit:

$$VaR_{t|t-1}(\alpha) = \text{percentile}(\{r_j\}_{j=t-Te}^{t-1}, 100\alpha) \quad (23)$$

Cette mesure de VaR non paramétrique est simple à mettre en oeuvre et pour cela très souvent utilisée en pratique. Elle repose sur l'hypothèse de distributions *i.i.d*, impliquant que la distribution conditionnelle de  $r_t$  soit invariante. Par conséquent, le fractile latent de la distribution conditionnelle est estimé par sa contrepartie empirique. Une telle méthode n'est certes pas exempte d'inconvénients, et de nombreux travaux empiriques, mais aussi théoriques ont souligné les insuffisances de la simulation historique (Hendricks, 1996; Boudoukh et al. 1998; Pritsker, 2001). La principale critique porte sur l'hypothèse de distribution *i.i.d*, car il est invraisemblable que les nombreux facteurs microstructurels et macroéconomiques concourant à la formation du prix d'un actif demeurent inchangés dans le temps. Cependant, comme le souligne Pritsker (2001), on peut toutefois s'accommoder du non respect de l'hypothèse de distributions *i.i.d*, en supposant que les observations les plus récentes de l'ensemble  $\Omega_{t-1}$  sont conditionnellement les plus importantes pour une éventuelle prévision. La pondération relative des éléments de l'ensemble  $\Omega_{t-1}$  que suggère ce raisonnement se retrouve chez Boudoukh et al. (1998) qui proposent une variante à la simulation historique, à savoir la méthode Hybride. L'approche demeure non paramétrique, et à l'instar de la simulation historique se résume à assimiler la VaR prévue au fractile empirique des rentabilités identifiées par l'ensemble  $\Omega_{t-1}$  retenu. L'avantage ici est qu'on exploite une information supplémentaire, à savoir le caractère plus informatif des rentabilités les plus proches de l'horizon de prévision. Techniquement, le calcul de la VaR Hybride se fait en trois étapes: (i) à chacune des  $Te$  rentabilités les plus

récentes  $r_{t-1}, r_{t-2}, \dots, r_{t-T_e}$  constituant l'ensemble  $\Omega_{t-1}$  est associée une pondération décroissante avec le temps de la forme  $\frac{1-\lambda}{1-\lambda^{T_e}}, \left(\frac{1-\lambda}{1-\lambda^{T_e}}\right) \lambda, \dots, \left(\frac{1-\lambda}{1-\lambda^{T_e}}\right) \lambda^{T_e-1}$  où  $\lambda < 1$ , puis (ii) les rentabilités (et les poids associés) sont ensuite ordonnées de manière croissante et (iii) enfin les poids ordonnés suivant les niveaux de rentabilité croissants sont sommés jusqu'à hauteur de  $\alpha\%$ . La VaR est alors égale à la rentabilité correspondant au dernier poids utilisé dans la sommation. Dans notre application, nous retiendrons des valeurs de  $\lambda$  et  $T_e$  respectivement fixées à 0.98 et 250.

Enfin, nous utiliserons une mesure semi-paramétrique de la VaR, à savoir le CAViaR proposée par Engle et Manganelli (2004). Dans ce cas, la modélisation porte directement sur le fractile latent, en lieu et place de l'ensemble de la distribution conditionnelle des rentabilités, comme pour les modèles précédents. Ainsi, la VaR à  $\alpha\%$  est défini par le modèle suivant :

$$VaR_{t|t-1}(\alpha) = \beta_0 + \beta_1 VaR_{t-1|t-2}(\alpha) + l(\beta_2, \dots, \beta_p, r_{t-1}, VaR_{t-1|t-2}(\alpha)) \quad (24)$$

où  $\beta_i \in \mathbb{R}$  et où  $l(\cdot)$  est une fonction de la rentabilité et de la VaR de la période précédente. Différentes spécifications sont proposées pour cette fonction. Nous retenons ici la spécification de type *Asymmetric Slope* :

$$l(\beta_2, \beta_3, r_{t-1}, VaR_{t-1|t-2}(\alpha)) = \beta_2 \max(r_{t-1}, 0) - \beta_3 \min(r_{t-1}, 0) \quad (25)$$

Les techniques de régressions quantiles non linéaires permettent d'estimer la valeur des paramètres  $\beta_i$  et ensuite celle de la VaR de la période courante. Le

modèle à estimer est alors le suivant<sup>5</sup> :

$$r_t = VaR_{t|t-1}(\alpha) + \varepsilon_t \quad Quant_\alpha(\varepsilon_t | \Omega_t) = 0 \quad (26)$$

La minimisation d'une fonction quantile objectif introduite par Koenker et Basset (1978) permet alors d'obtenir des estimateurs des paramètres  $\beta_i$  asymptotiquement convergents.

## 4 Evaluation des Procédures de Backtesting

La capacité des tests à discriminer entre "bonnes" et "mauvaises" mesures du risque est essentielle à toute procédure de validation de la VaR. Admettons que pour un portefeuille donné, on puisse obtenir  $N$  prévisions de la VaR à partir de modèles différents et non emboîtés. Supposons que ces  $N$  méthodes fournissent des prévisions significativement différentes non seulement en termes de VaR, mais aussi en termes d'occurrences des violations. Supposons enfin que l'on applique les procédures usuelles de backtesting à ces  $N$  prévisions. Admettons enfin, que les tests ne permettent pas de rejeter la validité pour les  $N$  prévisions de VaR (ou la plupart de ces prévisions). Deux interprétations peuvent alors être faites.

La première interprétation possible, à la manière de Candide, consisterait à postuler que ces différentes prévisions de VaR sont effectivement toutes valides. Il est en effet théoriquement possible que l'on soit en présence d'estimateurs sans biais du fractile de la distribution conditionnelle des rendements que les tests (notamment de couverture non conditionnelle) valideraient. La seconde interprétation consiste à penser que si les prévisions de VaR ne sont effectivement

---

<sup>5</sup>Remarquons la similitude avec le modèle linéaire traditionnel: la modélisation porte sur le fractile conditionnel (et non l'espérance conditionnelle), avec la nullité du fractile conditionnel du résidu (en lieu et place de l'espérance conditionnelle).

pas toutes valides, une telle configuration ne peut être que l'illustration de la faible puissance des tests de validation. Faible puissance qui se traduit par un non rejet de l'hypothèse nulle de validité et cela pour n'importe quelle prévision de risque. Cette seconde interprétation est, on le comprend aisément, beaucoup plus gênante pour les autorités de supervision ou les responsables du contrôle interne des risque.

Dans ce contexte, notre démarche d'évaluation des procédures de backtesting consiste à répliquer les tests de validation présentés dans la première section sur un grand nombre d'échantillons historiques de séries de rendements financiers pour lesquelles nous aurons établi des prévisions de VaR selon ces six méthodes.

Plus précisément, notre démarche est la suivante. On considère une série historique longue de rendements d'un actif financier à partir de laquelle on tire au hasard un sous échantillon de 500 points. A partir de ce sous échantillon, on estime (si nécessaire dans le cas des approches paramétriques et semi-paramétriques) les paramètres du modèle (modèle GARCH par exemple) censé représenter la distribution conditionnelle des rendements. A partir de ce modèle estimé (ou non), on construit une prévision hors période d'estimation de la VaR à 5% pour un horizon d'une période. En utilisant une fenêtre glissante, on reproduit cette démarche jusqu'à obtenir un échantillon de taille  $T = 250$  de prévisions  $VaR_{t|t-1}(\alpha)$ . On se situe ainsi exactement dans une perspective proche de la pratique en ne faisant porter les tests de validation que sur les prévisions hors période d'estimation (*out of sample*). A l'issue de cette étape, on dispose de six échantillons de prévisions de VaR à 5% associés aux six méthodes de calcul

retenues (Delta Normale, RiskMetrics, GARCH, simulation historique, hybride, CAViaR). Pour chacun de ces échantillons, on applique alors trois tests de couverture conditionnelle ( $LR_{CC}$ ,  $DQ_{CC}$ ,  $LR_{CC}^{duree}$ ), trois tests de l'hypothèse nulle d'indépendance des violations ( $LR_{IND}$ ,  $DQ_{IND}$ ,  $LR_{IND}^{duree}$ ) et un test de couverture non conditionnelle ( $LR_{UC}$ ) qui correspond à la statistique de Kupiec (1995). Pour un risque de première espèce (5% ou 10%), il est alors possible de déterminer suivant les tests retenus, combien parmi les six mesures alternatives de VaR peuvent être considérées comme valides. Plus précisément, on détermine le nombre de mesures pour lesquelles les tests ne permettent pas de rejeter l'hypothèse nulle de couverture non conditionnelle, d'indépendance ou de couverture conditionnelle.

Enfin, dans une démarche proche de celle adoptée pour l'étude des propriétés à distance finie des tests par simulations de Monte Carlo<sup>6</sup>, on tire au hasard un nouvel échantillon historique de rendements financiers et l'on reproduit l'ensemble de cette démarche. Dès lors, à partir de 500 expériences de ce type, il est possible d'observer les fréquences empiriques de validation conjointe des six mesures alternatives de VaR pour un même actif et d'évaluer ainsi la capacité des procédures de backtesting à discriminer les mesures de risques.

Cette procédure d'évaluation des procédures de backtesting est appliquée sur trois types d'actifs : l'indice Nasdaq, le titre General Motors et le taux de change Mark-Dollar US. L'échantillon total s'étend du 5 février 1971 au 16 août 2006 (soit 8 966 observations) pour l'indice Nasdaq et du 2 janvier 1962 au 16 août 2006 (soit 11 234 observations) pour le titre GM. Pour le taux de change Mark-

---

<sup>6</sup>La différence principale étant que, dans notre étude, ces échantillons tirés au hasard sont des échantillons historiques et non simulés selon un processus contrôlé : il nous est dès lors impossible de définir pour chaque échantillon la "vraie" distribution conditionnelle des pertes et profits et donc la "vraie" mesure de VaR.

Dollar US, la série considérée est la variation en pourcentage du taux spot mesurée toutes les dix minutes sur la période allant du 05 juin 1989 au 19 juin 1989, soit 2 488 observations.

## 5 Résultats

Dans cette section, sont présentés les résultats de notre procédure d'évaluation. Dans ce cadre, on vérifie tout d'abord que pour un échantillon donné, les six prévisions de VaR sont différentes et à tout le moins qu'il existe toujours au moins une mesure qui diffère sensiblement des cinq autres. A titre d'illustration sur la Figure 1 sont reportées les évolutions respectives des six prévisions *out-of-sample* de VaR à 5% sur le rendement de l'indice Nasdaq obtenues pour un échantillon tiré au hasard de 250 observations. Dans ce cas particulier, on observe que les prévisions de VaR divergent sensiblement : non seulement elles divergent entre grandes classes de méthodes (paramétriques et non paramétriques par exemple), mais aussi au sein d'une même approche, les mesures peuvent être relativement éloignées. De plus, les occurrences de violations, sur lesquelles sont fondées les tests de validation, n'interviennent pas nécessairement aux mêmes dates suivant la méthode de calcul retenue. Sur la Figure 2, sont ainsi reportées les valeurs cumulées, i.e.  $\sum_{s=1}^t I_s(\alpha)$ , des indicatrices associées aux violations des six prévisions de VaR. Ces valeurs ont été obtenues pour un autre échantillon de 250 observations du rendement associé à l'indice Nasdaq. Comme, on peut le constater, dans cet exemple précis, seules les méthodes delta normale et simulation historique conduisent exactement aux mêmes fonctions de répartition empirique des violations. Les autres VaR conduisent toutes à des processus de violations

ex-post qui surviennent à des dates sensiblement différentes. Ces constatations confirment les résultats qu'avait mis en évidence notamment Beder (1995) concernant la très grande sensibilité des mesures du risque au choix de la méthode de calcul de la VaR<sup>7</sup>.

Lorsque l'on considère pour un actif financier, l'ensemble des 500 échantillons tirés au hasard, il devient plus délicat de mettre en avant les différences existant entre les mesures de VaR. Toutefois dans le Tableau 1, nous proposons plusieurs indicateurs pour l'indice Nasdaq<sup>8</sup>, mettant en évidence les divergences dans les prévisions de la VaR et dans les violations observées *ex-post*. Tout d'abord, dans la partie haute du tableau sont proposés quatre statistiques mesurant les écarts entre d'une part les prévisions de VaR obtenus selon cinq méthodes (RiskMetrics, GARCH, simulation historique, hybride et CAViaR) et d'autre part la VaR Delta Normale, considérée comme référence. On observe ainsi que sur 500 répétitions, la moyenne des écarts minimums entre les différentes prévisions de VaR et la prévision delta normale varie de 0.003 (méthode CAViaR) à 0.037 (méthode Hybride). Rappelons, qu'en moyenne, la VaR Delta Normale à 5% définie pour l'indice Nasdaq est de l'ordre -0.014, ce qui implique qu'en moyenne, les écarts minimums représentent de 20% à 260% du niveau moyen de la VaR Delta Normale !

Dans la partie basse du tableau, sont reportées quatre statistiques permettant d'évaluer les différences observées entre les différentes séquences de violations ob-

---

<sup>7</sup>En utilisant huit mesures assez communes de VaR basées sur le croisement de trois critères (type de modélisation retenue pour la rentabilité du portefeuille, à savoir Simulation Historique ou Monte Carlo; hypothèse concernant les corrélations entre les actifs composant le portefeuille; hypothèse sur la période de détention), ce dernier avait en effet démontré qu'il existait de grandes différences entre les valeurs prévues de la VaR.

<sup>8</sup>Les résultats pour les deux autres actifs sont similaires et sont disponibles sur requête.

servées *ex-post*. Comme dans le cas précédent, la séquence des violations associées aux prévisions VaR Delta Normale est considérée comme référence. Parmi ces statistiques, figure la moyenne des écarts absolus entre les variables indicatrices de hit qui permet de rendre compte du nombre de violations qui n'apparaissent pas aux mêmes dates. On observe ainsi qu'il existe en moyenne sur 500 répliques, entre 10 et 11 violations associées aux prévisions GARCH qui ne coïncident pas avec les violations de prévisions VaR Delta Normale. Rappelons que pour une VaR à 5% sur un échantillon de 250 observations, on doit observer environ 12 violations. Les trois dernières mesures sont consacrées aux écarts existant entre les fonctions cumulatives des séquences de violations. Ces mesures permettent de rendre compte des écarts de "*timing*" entre la survenue d'une violation de la VaR calculée par une méthode donnée et la survenue d'une violation de la VaR calculée par la méthode delta normale. On observe ainsi que les moyennes des écarts sur ces cumulatives varient de  $-5.165$  pour la méthode GARCH à  $2.654$  pour les prévisions issues de la méthode hybride. L'écart minimum est obtenu pour la méthode de simulation historique comme c'était le cas dans l'échantillon d'exemple utilisé pour construire la Figure 2. Cela confirme que les violations apparaissent à des dates sensiblement différentes suivant la méthode de prévision retenue pour la VaR.

Toutefois, l'objectif de notre étude n'est pas simplement de poser le constat de ces différences entre les prévisions de risque : il s'agit de s'assurer que les procédures de backtesting et plus précisément les tests de validation permettent justement de discriminer parmi ces mesures alternatives un sous ensemble de



mesures valides et un sous ensemble de mesures non valides. Les résultats de notre procédure d'évaluation des tests usuels de backtesting pour les prévisions de VaR sur le rendement de l'indice Nasdaq sont reportés dans le Tableau 2. Pour chacun de sept tests considérés ( $LR_{CC}$ ,  $DQ_{CC}$ ,  $LR_{CC}^{duree}$ ,  $LR_{IND}$ ,  $DQ_{IND}$ ,  $LR_{IND}^{duree}$  et  $LR_{UC}$ ), sont reportées dans ce tableau les fréquences de non rejet de l'hypothèse nulle de validation pour six prévisions sur six de VaR et cela pour un risque de première espèce de 5% ou de 10%. Autrement dit, il s'agit de la fréquence avec laquelle les tests de backtesting conduiraient une personne en charge de la validation des prévisions de risque à ne rejeter aucune des six prévisions de VaR proposées. Les résultats obtenus à partir de 500 réplifications sont sans équivoque : quel que soit le risque de premier espèce retenu, quel que soit l'hypothèse testée (couverture conditionnelle, couverture non conditionnelle ou indépendance), quel que soit le test utilisé, les procédures de backtesting ont une tendance très forte à conclure que *"tout va pour le mieux dans le meilleur des mondes"*.

Considérons par exemple le test de couverture conditionnel (noté  $LR_{CC}$ ) proposé par Christoffersen (1998) : dans 24% des expériences, pour un seuil de risque de 5%, ce test conduit à ne pas rejeter la validité d'une seule des six prévisions de VaR proposées. Dis autrement, à partir de données de prévisions portant sur une année complète de cotations (250 jours), dans plus d'un cas sur cinq, un utilisateur de ce test serait donc conduit, dans une sorte d'optimisme leibnizien, à ne pas rejeter la validité de toutes les prévisions de VaR que l'on lui propose. Ces résultats sont encore plus nets lorsque l'on considère les fréquences avec lesquelles les tests permettent de rejeter la validité d'au plus une mesure de VaR sur les six

étudiées. On observe en effet que dans près d'un cas sur deux, le test  $LR_{CC}$  ne permet pas de rejeter la validité pour au moins cinq des six prévisions de VaR étudiées.

Ces résultats ne sont pas propres aux tests LR de Christoffersen. Tous les tests usuels de backtesting conduisent, dans des proportions variables, à des conclusions similaires. On note toutefois que les tests  $DQ$  d'Engle et Manganelli (2004) semblent nettement plus restrictifs : pour le test de couverture conditionnelle, cinq prévisions sur six ont été validées dans seulement 14% des expériences. Mais cette fréquence passe tout de même à 23%, dès que lors que l'on considère de validation d'au moins quatre prévisions de VaR. Ces résultats confirment les expériences standards d'évaluation des propriétés à distance finie des tests de Backtesting (Berkowitz, Christoffersen et Pelletier, 2005 ; Hurlin et Tokpavi, 2007) dans lesquels généralement on montre que le test  $DQ$  est l'un des plus puissants à distance finie.

Enfin, ces observations sont globalement identiques pour les mesures de VaR calculées sur le rendement du titre General Motors (Tableau 3) et sur le taux de change Mark Dollar (Tableau 4). La tendance à la validation globale de n'importe quelle prévision de VaR est encore plus marquée dans le cas des prévisions de VaR sur le rendement du titre GM. Par exemple, si l'on considère le test  $DQ_{CC}$  de couverture conditionnelle d'Engle et Manganelli (2004), la probabilité qu'au moins cinq prévisions de VaR sur six sont considérées comme valide est de 50%. Dans 20% des échantillons, ce test ne permet pas de rejeter la validité pour ne serait qu'une seule prévision de VaR. En ce qui concerne, le rendement sur le taux de change marks – dollars, notons qu'une personne appliquant une procédure de

backtesting fondée sur le test de Christoffersen (1998) aura trois chances sur six de valider au moins cinq prévisions alternatives de VaR sur six. Autant dire, que cette procédure semble à tout le moins peu discriminante.

Sur le plan technique, ces résultats ne sont que la conséquence de la faible puissance des tests de validation sur petits échantillons. Il est toujours difficile d'établir un diagnostic général sur les propriétés à distance finie de ces tests, puisque qu'en particulier leur puissance dépend de "l'écart" entre la vraie VaR et la méthode de calcul utilisée. Toutefois, les différents exercices proposés dans la littérature (Campbell, 2007; Berkowitz, Christoffersen et Pelletier, 2005; Hurlin et Tokpavi, 2007; etc..) montrent que ces tests présentent des tailles empiriques proches de leur niveau nominal et des puissances relativement faibles. Par exemple, pour une VaR à 5% et un risque nominal de 10%, le test de couverture conditionnelle  $LR_{CC}$  de Christoffersen (1998) présente une puissance de l'ordre de 30% sous différentes alternatives pour un échantillon de 250 observations. Le test  $DQ_{CC}$  d'Engle et Manganelli (2004) permet d'atteindre une puissance de l'ordre de 40%. Un test fondé sur une statistique de Portemanteau multivariée (Hurlin et Tokpavi, 2007) permet d'atteindre des puissances supérieures à 50% et cela pour différentes configurations d'exercices de simulations. Mais globalement, on peut affirmer que pour un échantillon de 250 observations, lorsque la méthode de calcul de la VaR est non valide, les tests permettent de rejeter l'hypothèse de couverture conditionnelle, et donc la validité de la VaR, dans environ un cas sur trois, voir un cas sur deux pour les tests les plus puissants. Notre approche d'évaluation met en lumière le fait que cette puissance, somme toute raisonnable, implique pourtant que pour un même actif, les procédures actuelles de backtest-

ing ont dans l'ensemble tendance à ne pas rejeter la validité d'un grand nombre de mesures alternatives du risque pourtant sensiblement différentes.

## 6 Conclusion

Les résultats de notre étude montrent que les procédures actuelles de Backtesting sont très peu discriminantes. Pour les trois types d'actifs considérés, il s'avère que les tests usuels menés à partir d'échantillon de 250 observations (soit environ une année de cotations), ont une forte tendance à ne pas rejeter la validité de la plupart des prévisions (paramétriques ou non paramétriques) de VaR considérées. Ainsi, "*tout va pour le mieux dans le meilleur des mondes*", serait tenté de conclure le candide qui, selon ces procédures, validerait la plupart des prévisions de VaR qui lui seraient présentées.

Une telle analyse de nos résultats donnerait en outre des arguments en faveur de ceux qui privilégieraient des méthodes « simples » de calcul de la VaR (par exemple issue d'un modèle GARCH estimée sur le rendement agrégé d'un portefeuille) au détriment de méthodes plus élaborées tenant compte par exemple des covariances entre les rendements des différents actifs (modèles GARCH multivariés, modèles DCC, etc.). En effet, si les procédures de Backtesting ne permettent généralement pas de rejeter la validité pour la plupart des mesures étudiées, est-il vraiment nécessaire, par exemple, de modéliser les corrélations dynamiques du portefeuille pour prévoir la VaR ?

Les esprits Voltairiens analyseront nos résultats d'une toute autre façon. Ces résultats confirment l'intérêt qu'il y a à développer de nouvelles procédures de tests de validation, à la fois simples et plus puissantes. Simple, dans la mesure

où le principe des tests et leurs résultats doivent pouvoir être communiqués à des non spécialistes et appliqués à toute prévision issue de n'importe quel modèle (approche de type *model free*). Puissante, dans la mesure où ces procédures devront permettre de rejeter plus fréquemment la validité des prévisions non valides à partir de petits échantillons.

## A Bibliographie

- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. AND LABYS, P. (2003), "Modeling and Forecasting Realized Volatility", *Econometrica*, 71, p. 579-625.
- BEDER, T. (1995), "VaR: Seductive but Dangerous", *Financial Analysts Journal* 51, 5, p. 12-24.
- BERKOWITZ, J. (2001), "Testing Density Forecasts With Applications to Risk Management", *Journal of Business and Economic Statistics* 19, p. 465-474.
- BERKOWITZ, J., ET J. O BRIEN. (2002), "How Accurate are the Value-at-Risk Models at Commercial Banks", *Journal of Finance* 57, p. 1093-1111.
- BERKOWITZ, J, CHRISTOFFERSEN P.F. ET PELLETIER (2005), "Evaluating Value-at-Risk Models with Desk-Level Data", *North Carolina State University, Department of Economics, Working Paper* 010.
- BOUDOUGH J, M RICHARDSON ET R WHITELAW, (1998), "The Best of Both Worlds", *Risk*, 11, p. 64-67.
- CAMPBELL, S. D. (2007), "A Review of Backtesting and Backtesting Procedures", *Journal of Risk*, Vol 9, Number 2, p. 1-19.
- CHRISTOFFERSEN, P. F.. (1998), "Evaluating interval forecasts", *International Economic Review* 39, p. 841-862.
- CHRISTOFFERSEN, P. F., ET D. PELLETIER. (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics* 2, 1, p. 84-108.
- DIEBOLD, F. X. ET MARIANO, R. S. (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13, p. 253-263.
- DOWD, K. (2005), *Measuring market risk*, John Wiley & Sons Ltd.
- ENGLE, R. ET BOLLERSLEV, T., (1986), "Modeling the Persistence of Conditional Variances", *Econometric Review*, Vol. 5., p. 1-50.
- ENGLE, R. F., ET S. MANGANELLI. (2004), "CAViaR: Conditional autoregressive Value-at-Risk by regression quantiles", *Journal of Business and Economic Statistics* 22, p. 367-381.
- HENDRICKS D., (1996), "Evaluation of Value-at-Risk Models Using Historical Data", *Federal Reserve Bank of New York Economic Policy Review*, April, p. 39-69.
- HURLIN C. AND TOKPAVI, S. (2007), "Bactesting Value-at-Risk Accuracy: A Simple New Test", *Journal of Risk*, Vol 9, 2, p. 19-37.
- KOENKER R. AND BASSETT G. (1978), "Regression Quantiles", *Econometrica*, 46, p. 33-50.

KUPIEC, P.. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives* 3, p. 73-84.

PRITSKER M., (2001), "The Hidden Dangers of Historical Simulation", *Finance and Economics Discussion Series* 2001-27, Board of Governors of the Federal Reserve System.

Figure 1: Prévisions *out-of-sample* de VaR à 5%, Indice Nasdaq,  $T = 250$

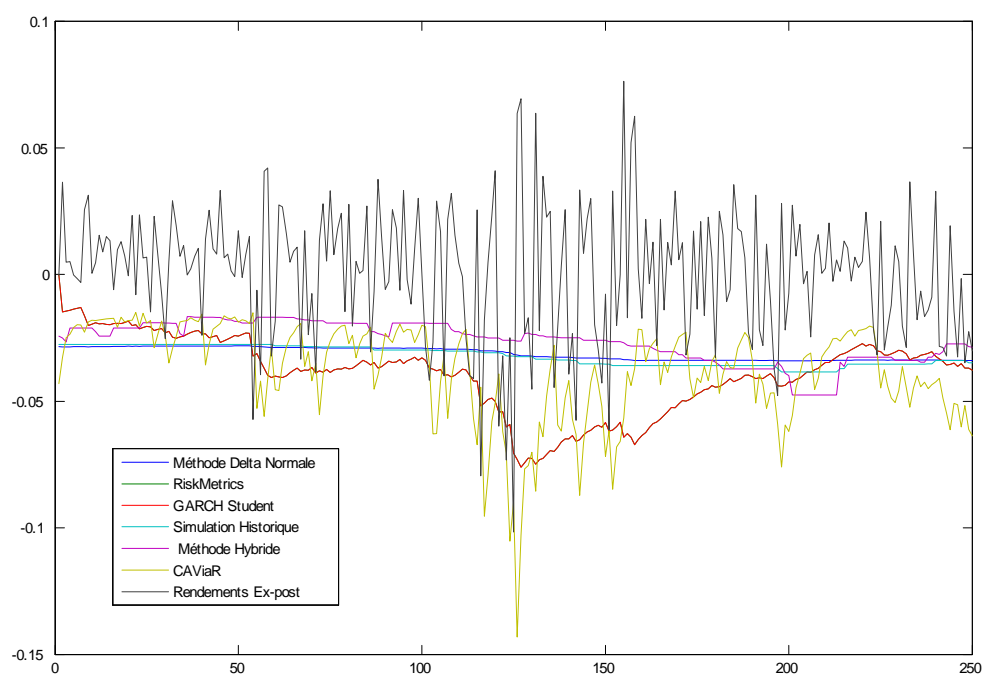


Figure 2: Sommes Cumulées des Fonctions Indicatrices de Violations  $I_t(\alpha)$ , VaR à 5%, Indice Nasdaq,  $T = 250$

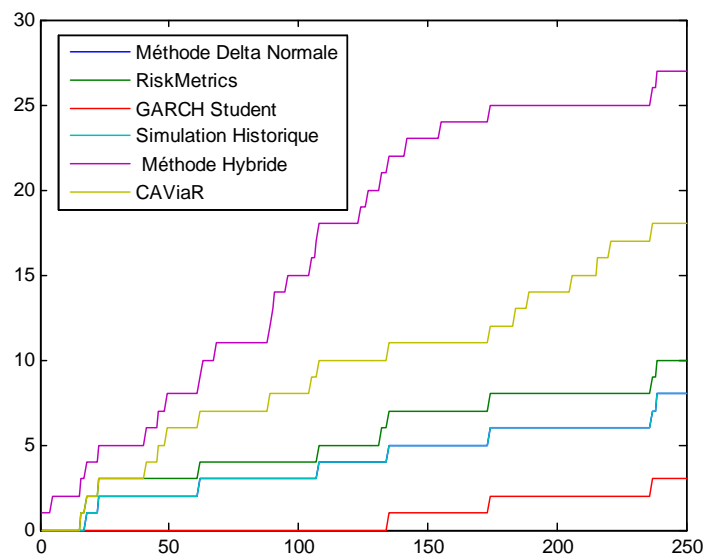




Tableau 1: Mesures de Distances entre les VaR, Indice Nasdaq<sup>9</sup>

<i>Mesures de VaR 5% - Indice Nasdaq</i>					
Indicateur	RiskMetrics	GARCH	H.S.	Hybride	CAViaR
$\frac{1}{S} \sum_{s=1}^S \max_{\{t\}} \left\{ VaR_t^{(s)} - \tilde{VaR}_t^{(s)} \right\}$	1.646	2.632	0.204	0.914	3.967
$\frac{1}{S} \sum_{s=1}^S \min_{\{t\}} \left\{ VaR_t^{(s)} - \tilde{VaR}_t^{(s)} \right\}$	0.004	0.004	0.031	0.037	0.003
$\frac{1}{S \times T} \sum_{s=1}^S \sum_{t=1}^T \left[ VaR_t^{(s)} - \tilde{VaR}_t^{(s)} \right]$	-0.123	-0.169	-0.031	0.192	-0.028
$\frac{1}{S \times T} \sum_{s=1}^S \sum_{t=1}^T \left[ VaR_t^{(s)} - \tilde{VaR}_t^{(s)} \right]^2$	0.005	0.005	0.0002	0.003	0.095
<i>Variable Indicatrice associée aux Violations de la VaR (Hit Function)</i>					
Indicateur	RiskMetrics	GARCH	H.S.	Hybride	CAViaR
$\frac{1}{S} \sum_{s=1}^S \sum_{t=1}^T \left  I_t^{(s)} - \tilde{I}_t^{(s)} \right $	8.858	10.64	2.624	9.952	12.93
$\frac{1}{S} \sum_{s=1}^S \max_{\{t\}} \left\{ \sum_{j=1}^t \left[ I_j^{(s)} - \tilde{I}_j^{(s)} \right] \right\}$	2.236	0.180	0.716	7.258	4.192
$\frac{1}{S} \sum_{s=1}^S \min_{\{t\}} \left\{ \sum_{j=1}^t \left[ I_j^{(s)} - \tilde{I}_j^{(s)} \right] \right\}$	-4.190	-9.926	-1.760	-1.774	-3.124
$\frac{1}{S \times T} \sum_{s=1}^S \sum_{t=1}^T \sum_{j=1}^t \left[ I_j^{(s)} - \tilde{I}_j^{(s)} \right]$	-1.161	-5.165	-0.486	2.654	0.524

Notes :  $VaR_t^{(s)}$  désigne la mesure de VaR (RiskMetrics, GARCH, HS, Hybride, CAViaR) obtenue la date  $t$  dans le  $s^{\text{ème}}$  échantillon ( $s = 1, \dots, S$ ) et  $\tilde{VaR}_t^{(s)}$  la mesure de référence associée à la méthode Delta Normale. L'indicatrice  $I_t^{(s)}$  (ou *Hit Function*) prend la valeur 1 si le rendement observé est inférieur à la VaR et 0 sinon. L'indicatrice  $\tilde{I}_t^{(s)}$  est associée à la mesure de référence  $\tilde{VaR}_t^{(s)}$ .

Tableau 2: Validation d'une VaR 5%, indice nasdaq,  $T = 250^{10}$

<i>Tests de Christoffersen (1998) et de Kupiec (1995)</i>			
Seuil Nominal 5%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.244	0.476	0.046
Fréquence de rejet d'au plus 1 mesure de VaR	0.450	0.668	0.292
Fréquence de rejet d'au plus 2 mesures de VaR	0.564	0.758	0.468
Seuil Nominal 10%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.190	0.416	0.006
Fréquence de rejet d'au plus 1 mesure de VaR	0.372	0.552	0.208
Fréquence de rejet d'au plus 2 mesures de VaR	0.510	0.642	0.410
<i>Tests de Engle et Manganelli (2004)</i>			
Seuil Nominal 5%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 6 mesures de VaR	0.082	0.146	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.224	0.318	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.338	0.492	—
Seuil Nominal 10%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 4 mesures de VaR	0.028	0.100	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.146	0.272	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.238	0.386	—
<i>Tests de Christoffersen et Pelletier (2004)</i>			
Seuil Nominal 5%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.076	0.580	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.366	0.792	—
Fréquence de rejet d'au plus de VaR	0.508	0.870	—
Seuil Nominal 10%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.022	0.424	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.258	0.690	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.438	0.814	—

Notes : Les tests de l'hypothèse nulle de couverture non conditionnelle ( $LR_{UC}$ ), d'indépendance des violations ( $LR_{IND}$ ,  $DQ_{IND}$ ,  $LR_{IND}^{duree}$ ) ou de couverture conditionnelle ( $LR_{CC}$ ,  $DQ_{CC}$ ,  $LR_{CC}^{duree}$ ) sont appliqués sur six mesures de VaR à 5% (Delta Normale, Risk-Metrics, GARCH, HS, Hybride, CAViaR) à partir de 500 échantillons de taille  $T = 250$ .

Tableau 3: Validation d'une VaR 5%, titre GM,  $T = 250^{11}$

<i>Tests de Christoffersen (1998) et de Kupiec (1995)</i>			
Seuil Nominal 5%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.478	0.824	0.146
Fréquence de rejet d'au plus 1 mesure de VaR	0.714	0.944	0.474
Fréquence de rejet d'au plus 2 mesures de VaR	0.812	0.954	0.664
Seuil Nominal 10%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.396	0.732	0.102
Fréquence de rejet d'au plus 1 mesure de VaR	0.662	0.906	0.368
Fréquence de rejet d'au plus 2 mesures de VaR	0.78	0.942	0.532
<i>Tests de Engle et Manganelli (2004)</i>			
Seuil Nominal 5%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 6 mesures de VaR	0.204	0.290	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.518	0.664	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.708	0.806	—
Seuil Nominal 10%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 4 mesures de VaR	0.102	0.208	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.328	0.522	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.576	0.692	—
<i>Tests de Christoffersen et Pelletier (2004)</i>			
Seuil Nominal 5%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.186	0.748	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.488	0.878	—
Fréquence de rejet d'au plus de VaR	0.696	0.952	—
Seuil Nominal 10%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.124	0.602	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.380	0.770	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.578	0.866	—

Notes : Les tests de l'hypothèse nulle de couverture non conditionnelle ( $LR_{UC}$ ), d'indépendance des violations ( $LR_{IND}$ ,  $DQ_{IND}$ ,  $LR_{IND}^{duree}$ ) ou de couverture conditionnelle ( $LR_{CC}$ ,  $DQ_{CC}$ ,  $LR_{CC}^{duree}$ ) sont appliqués sur six mesures de VaR à 5% (Delta Normale, Risk-Metrics, GARCH, HS, Hybride, CAViaR) à partir de 500 échantillons de taille  $T = 250$ .

Tableau 4: Validation d'une VaR 5%, Mark-Dollars,  $T = 250^{12}$

<i>Tests de Christoffersen (1998) et de Kupiec (1995)</i>			
Seuil Nominal 5%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.472	0.66	0.010
Fréquence de rejet d'au plus 1 mesure de VaR	0.752	0.818	0.246
Fréquence de rejet d'au plus 2 mesures de VaR	0.960	1.000	0.642
Seuil Nominal 10%	$LR_{CC}$	$LR_{IND}$	$LR_{UC}$
Fréquence de validitation des 6 mesures de VaR	0.424	0.632	0.000
Fréquence de rejet d'au plus 1 mesure de VaR	0.700	0.784	0.214
Fréquence de rejet d'au plus 2 mesures de VaR	0.940	1.000	0.478
<i>Tests de Engle et Manganelli (2004)</i>			
Seuil Nominal 5%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 6 mesures de VaR	0.160	0.432	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.416	0.530	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.812	0.846	—
Seuil Nominal 10%	$DQ_{CC}$	$DQ_{IND}$	—
Fréquence de validitation des 4 mesures de VaR	0.060	0.380	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.282	0.486	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.730	0.798	—
<i>Tests de Christoffersen et Pelletier (2004)</i>			
Seuil Nominal 5%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.012	0.762	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.494	1.000	—
Fréquence de rejet d'au plus de VaR	0.776	1.000	—
Seuil Nominal 10%	$LR_{CC}^{duree}$	$LR_{IND}^{duree}$	—
Fréquence de validitation des 6 mesures de VaR	0.000	0.590	—
Fréquence de rejet d'au plus 1 mesure de VaR	0.306	0.818	—
Fréquence de rejet d'au plus 2 mesures de VaR	0.662	0.980	—

Notes : Les tests de l'hypothèse nulle de couverture non conditionnelle ( $LR_{UC}$ ), d'indépendance des violations ( $LR_{IND}$ ,  $DQ_{IND}$ ,  $LR_{IND}^{duree}$ ) ou de couverture conditionnelle ( $LR_{CC}$ ,  $DQ_{CC}$ ,  $LR_{CC}^{duree}$ ) sont appliqués sur six mesures de VaR à 5% (Delta Normale, Risk-Metrics, GARCH, HS, Hybride, CAViaR) à partir de 500 échantillons de taille  $T = 250$ .