



HAL
open science

Quelle régulation pour les hôpitaux publics français ?

Brigitte Dormont, Carine Milcent

► **To cite this version:**

Brigitte Dormont, Carine Milcent. Quelle régulation pour les hôpitaux publics français?. Revue Française d'Economie, 2002, 17 (2), pp.117-142. 10.3406/rfec0.2002.1463 . halshs-01990530

HAL Id: halshs-01990530

<https://shs.hal.science/halshs-01990530>

Submitted on 23 Jan 2019

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.

Quelle régulation pour les hôpitaux publics français ?

Brigitte Dormont, Carine Milcent

Citer ce document / Cite this document :

Dormont Brigitte, Milcent Carine. Quelle régulation pour les hôpitaux publics français ?. In: Revue française d'économie, volume 17, n°2, 2002. pp. 117-142;

doi : <https://doi.org/10.3406/rfeco.2002.1463>

https://www.persee.fr/doc/rfeco_0769-0479_2002_num_17_2_1463

Fichier pdf généré le 21/04/2018

Abstract

The purpose of this paper is to study hospital costs in the event of introduction of a Prospective Payment System in France. We use a nested three dimensional database (stays- hospitals-years) in order to identify hospital unobservable heterogeneity and a transitory moral hazard component of cost variability. Econometric estimates are performed on a sample of 7,314 stays for acute myocardial infarction observed in 36 French public hospitals over the period 1994 to 1997. Transitory moral hazard is far from negligible: its estimated standard error is about 50 % of the standard error we estimate for cost variability due to permanent unobservable heterogeneity between hospitals. Simulations show that a cost reduction of about 16 % can be expected from implementation of a payment system which allows for permanent unobserved heterogeneity and eliminates only transitory moral hazard.

Résumé

Cet article étudie les coûts des hôpitaux publics français dans la perspective d'une réforme de la tarification. Nous utilisons des données à trois dimensions hôpital-séjour-année afin d'identifier l'hétérogénéité non observée des hôpitaux et la composante de la variabilité du coût attribuable à l'aléa moral transitoire. Les applications empiriques sont réalisées sur un échantillon de 7 314 séjours pour infarctus du myocarde aigu observés dans 36 hôpitaux publics français sur la période 1994-1997. L'aléa moral transitoire est loin d'être négligeable. Son écart-type estimé s'élève à 50 % de la variabilité des coûts due à l'hétérogénéité non observée des hôpitaux. Les simulations montrent que l'on peut attendre des économies budgétaires de 16 % environ de l'application d'une tarification qui prend en compte les hétérogénéités observées entre les hôpitaux et n'élimine que l'aléa moral transitoire.

**Brigitte
DORMONT
Carine
MILCENT**

Quelle régulation pour les hôpitaux publics français ?

T

ous les acteurs du système hospitalier se plaignent de la pénurie des moyens à leur disposition. Dans le même temps, un consensus se dégage pour constater les inégalités de dotations budgétaires et les inefficacités qui caractérisent l'activité hospitalière (Mougeot [1999]). Une réforme de la tarification est à l'ordre du jour, avec l'expérimentation d'un système de tarification par pathologie.

Cet article a pour objet d'étudier la variabilité des coûts des hôpitaux publics français dans la perspective d'une réforme de leur tarification. Les applications empiriques sont réalisées sur des données à trois dimensions hôpital-séjour-année concernant des patients traités pour infarctus du myocarde aigu. Notre approche consiste à essayer d'identifier les inefficacités repérables afin de définir des formules de paiement correspondant à des coûts de traitements efficaces. L'application d'une telle tarification est alors simulée sur les données afin d'évaluer les gains budgétaires qui pourraient en découler.

Dans cet article, une mise en perspective historique sur la régulation des hôpitaux précède une présentation de la logique empirique de l'évaluation que nous proposons, sur données françaises. Puis l'analyse économétrique de la variabilité des coûts hospitaliers permet de déboucher sur les simulations de tarification. Enfin, nous concluons sur les avantages de la tarification que nous proposons et la portée de l'analyse réalisée.

La régulation des hôpitaux : mise en perspective historique et éléments de théorie

Le secteur hospitalier doit faire l'objet d'une régulation. Cet impératif ne découle pas du statut des établissements, qui peut être privé, public ou privé à but non lucratif. Il est lié au fait que la demande qui s'adresse aux hôpitaux, solvabilisée par l'assurance-maladie, est peu influencée par les mécanismes de marché. Les établissements hospitaliers peuvent être régulés par des tutelles aux statuts forts différents. Il peut s'agir de l'Etat (hôpitaux publics français), d'un assureur universel obligatoire (le programme Medicare pour les retraités américains) ou

encore d'un assureur privé (cas des HMO aux Etats-Unis). Notre étude s'intéresse aux hôpitaux publics français.

Le régulateur doit encourager les hôpitaux à répondre aux besoins, sans induire d'activité injustifiée. Il doit inciter les établissements à l'efficacité, c'est à dire à fournir les soins au moindre coût, sans sacrifier la qualité ni sélectionner les patients. Enfin, le régulateur doit distribuer les ressources de manière équitable entre les hôpitaux et, dans cette perspective, être capable d'évaluer l'activité de production de soins des établissements et les contraintes spécifiques auxquelles certains d'entre eux peuvent être confrontés.

Plusieurs types de régulation peuvent être adoptés. Le paiement à l'acte finance de manière rétrospective les soins au niveau du coût effectivement observé. Le système du budget global consiste à attribuer à l'établissement, selon des critères qui peuvent varier, un budget pour l'ensemble de son activité annuelle. La tarification par pathologie consiste à financer chaque cas traité par un forfait calculé de façon prospective pour chaque pathologie. Le principe de l'achat de soins, enfin, consiste à sélectionner par un mécanisme d'enchères le prestataire d'un panier de soins prédéfinis.

Dans ce qui suit, nous faisons une brève présentation des évolutions récentes de la régulation des hôpitaux en France et aux Etats-Unis dans le système Medicare. Cet exemple est convoqué à titre d'expérience du mode de régulation qui pourrait être envisagé pour la France. Puis nous donnons des repères théoriques permettant de comprendre les principes de la tarification par pathologie et son évolution vers un concept de paiement mixte.

Le budget global en France : inefficacités et iniquité

Comme les hôpitaux de la plupart des pays de l'OCDE, les hôpitaux français ont tout d'abord été financés par un système de paiement rétrospectif. Ce système permet un haut niveau de qualité des soins mais n'incite pas à l'efficacité.

Entre 1945 et 1983, le paiement rétrospectif prend la forme du calcul d'un prix de journée qui fonde la procédure d'allocation des ressources. L'absence de contrainte budgétaire qui prévaut durant cette période se traduit par une forte dérive des dépenses hospitalières mais aussi par le développement d'une médecine de pointe (Moison et Tonneau [1997]).

En 1983 est instauré le système du budget global. De façon générale, un tel système permet de contrôler la dépense hospitalière, mais risque d'entraîner une diminution du volume de services rendus et de la qualité des soins.

Dans le cas français, le budget est établi par l'application d'un taux directeur à la dotation de l'exercice précédent. Ce dispositif a contribué à figer les inégalités de dotations initiales indépendamment de l'évolution de l'activité des hôpitaux.

Concrètement, la sévérité de la contrainte budgétaire subie par un hôpital dépend alors de trois éléments : (i) le niveau d'efficacité de l'établissement lors de l'année servant de base historique au calcul du budget global ; (ii) son dynamisme en matière d'activité et d'acquisition des innovations techniques ; (iii) le pouvoir de négociation du directeur de l'hôpital auprès de la tutelle afin d'influencer le budget alloué. Il est clair que l'efficacité initiale et le dynamisme influencent positivement le niveau et la qualité du service rendu : dans ce système, ils jouent dans le sens d'un resserrement de la contrainte budgétaire, toutes choses égales par ailleurs. Le système du budget global, tel qu'il est appliqué actuellement, permet ainsi des inefficacités dans l'activité des établissements et conduit à une allocation des ressources inéquitable.

Conscientes de ces imperfections, les autorités de régulation ont très vite cherché à mettre en place une collecte d'informations sur l'activité de production de soins des établissements. Le lancement en 1982 du Programme de médicalisation des systèmes d'informations (PMSI) répond à cet objectif. Opérationnel à partir de 1994 seulement, le PMSI permet d'évaluer les services rendus par les hôpitaux en nombre de points ISA (indice synthétique d'activité). Les ordonnances

de 1996 réforment le système hospitalier en créant les agences régionales de l'hospitalisation (ARH), tutelles uniques au niveau régional. Les directeurs des ARH sont dotés d'une position statutaire qui leur garantit en principe une indépendance vis-à-vis des pressions de diverses natures. Ils ont toute latitude pour appliquer une logique d'égalisation des dotations en effectuant les arbitrages budgétaires sur la base du nombre de points ISA réalisés par les établissements.

Le bilan des ordonnances de 1996 est mitigé : les redéploiements de ressources n'ont pas eu lieu, peut-être à cause d'un contexte difficile de restrictions budgétaires. Le budget global est toujours en place, avec les inefficacités et les allocations inéquitables afférentes. La tutelle examine maintenant sérieusement l'opportunité de se tourner vers un autre mode de régulation : la tarification à la pathologie, dont une expérimentation est prévue dans l'article L716.2 du code de la santé publique.

L'expérience de Medicare aux Etats-Unis : de multiples amendements à la tarification forfaitaire

Système d'assurance-maladie des retraités américains de 65 ans et plus, le programme Medicare rompt dès 1983 avec le paiement rétrospectif en instaurant un système de tarification forfaitaire pour rembourser les séjours hospitaliers de ses ayants droit. Le but recherché est de parvenir à une maîtrise des dépenses hospitalières.

Très rapidement, on s'est écarté de l'application stricte de paiements prospectifs (Hartmann *et alii*, [2001]). Des ajustements ont été introduits pour tenir compte des disparités des coûts salariaux sur le territoire américain, des différences de statut entre les établissements (mission d'enseignement) ou encore de l'existence d'une forte proportion de patients à bas revenus ou souffrant de maladies chroniques. Des remboursements additionnels ont été créés pour aider les hôpitaux à supporter les coûts dans les cas extrêmes, par exemple dans le cas d'une durée de séjour particulièrement longue. Avec la mise en place

d'instances de contrôle, ces nombreux aménagements au système initial ont été introduits afin de limiter les comportements déviants auxquels peut inciter un paiement forfaitaire trop rigide : sélection de clientèle, baisse de la qualité des soins ou encore manipulation du codage des séjours (pour aboutir à une pathologie mieux rémunérée).

Ainsi, la tarification a progressivement intégré une proportion croissante de paiements à caractère rétrospectif (McClellan, [1997]). Cette évolution peut être invoquée comme explication à l'efficacité apparemment limitée de la tarification à la pathologie sur la maîtrise des coûts aux Etats-Unis.

Principe de la tarification par pathologie et évolution vers un concept de paiement mixte

Le pragmatisme de l'administration de Medicare l'a conduite à introduire ces multiples aménagements pour assouplir le système initial. Dans le même temps, de nombreuses contributions théoriques faisaient évoluer le modèle de base en donnant des fondements théoriques à des paiements mixtes, combinant forfait et remboursements rétrospectifs.

La modélisation retenue suppose en général que le coût de traitement d'un patient atteint d'une pathologie particulière dans l'hôpital h est donné par : $C_h = c_h - e_h$, c_h et e_h sont des informations privées de l'hôpital ; c_h est un paramètre qui reflète sa productivité ; e_h est l'effort de réduction du coût, effort qui implique une désutilité pour l'hôpital. L'objectif du régulateur est d'inciter l'hôpital à l'efficacité et donc à fournir le niveau d'effort maximal.

La tarification par pathologie consiste à proposer un paiement forfaitaire pour chaque cas traité dans une pathologie donnée. Le forfait proposé étant défini indépendamment de son coût, l'hôpital a intérêt à réduire ses coûts et donc à fournir l'effort optimal de premier rang. A ce stade, une partie seulement du problème est résolue. En effet, c_h est une information privée

de l'hôpital : la définition du forfait par la tutelle peut le conduire à la faillite ou lui attribuer des rentes excessives.

Le problème de la tutelle est alors de trouver le niveau de paiement correspondant à la production efficace, autrement dit de définir une formule de tarification permettant d'extraire la totalité – ou de réduire le plus possible – la rente informationnelle.

Ce problème informationnel peut être résolu en retenant une hypothèse simplificatrice d'homogénéité des hôpitaux : $c_h = c \forall h$ (Shleifer [1985]). Dans ce cas les disparités de coûts sont exclusivement dues à l'aléa moral : $C_h = c - e_h$, et la tutelle peut mettre en place un mécanisme de concurrence par comparaison. Celui-ci consiste à proposer à chaque hôpital un paiement défini sur la base de la moyenne des coûts observés pour les autres établissements en fin d'exercice. On montre alors que cette règle de paiement conduit à la fixation d'un forfait qui correspond au coût de production efficace.

Cette représentation idéale constitue le fondement théorique de la tarification par pathologie. Elle repose sur des hypothèses peu réalistes : homogénéité des hôpitaux, homogénéité des patients pour une pathologie donnée, qualité des soins fixée. Elle conduit à définir une tarification au forfait, trop homogène, dont les risques sont maintenant connus : sélection ou discrimination des patients, baisse de la qualité des soins délivrés (Newhouse [1996]).

Les contributions théoriques ont cherché à améliorer le modèle de base en remettant en cause l'hypothèse d'homogénéité des patients (Keeler [1990], Pope [1990], Ma [1994, 1998], Ellis [1998]), ou en remettant en cause l'hypothèse d'homogénéité des hôpitaux (Auriol et Laffont [1992], Laffont et Tirole [1993], Ellis [1998]). A partir d'hypothèses et d'approches théoriques très différentes, ces modèles conduisent à définir des tarifications optimales correspondant à des paiements mixtes, combinant forfait et coûts observés.

Le principe d'un système de paiement mixte fait maintenant l'objet d'un consensus, rejoignant ainsi la pratique de la régulation aux Etats-Unis. Mais le coefficient de partage entre

forfait et coût effectif est défini de façon très différente selon le modèle théorique auquel on se réfère, son corps d'hypothèses et sa paramétrisation. De plus, ce coefficient peut dépendre de variables ou de fonctions non observables (comme par exemple la désutilité de l'effort chez Laffont et Tirole [1993]).

Le problème débouche alors sur des questions essentiellement empiriques que cet article aborde dans le cas français : comment identifier le niveau des coûts correspondant à une activité efficace ? Comment intégrer l'hétérogénéité des patients et celle des établissements dans la tarification des séjours hospitaliers ?

Logique empirique d'une évaluation sur données françaises

La situation française est particulière dans la mesure où un système d'information relativement sophistiqué sur l'activité hospitalière a déjà été mis en place, sans conséquence notable sur la procédure de détermination des budgets. De ce fait, les évaluations proposées ne peuvent pas reposer sur l'observation des effets d'un changement de tarification intervenu globalement, ou à titre d'expérience pour un ensemble restreint d'établissements.

Nos observations des coûts se situent au niveau individuel du séjour hospitalier. Ces coûts résultent d'une activité financée selon le système du budget global. Ils reflètent donc en partie le mode de fixation du budget : application d'un taux directeur au budget précédent, négociation plus ou moins fructueuse du gestionnaire de l'hôpital auprès de la tutelle.

Sur le terrain, les établissements disposant d'une bonne comptabilité évaluent les coûts par séjour et transmettent cette information en vue de la constitution de la base de coût PMSI, que nous utilisons pour les évaluations empiriques. Dans ce cadre, la variabilité des coûts observés par séjour est influencée par plusieurs éléments : (i) les caractéristiques des patients

traités qui impliquent effectivement des coûts de traitement différenciés, (ii) les caractéristiques des établissements (infrastructures, économies de gamme ou d'échelle), (iii) les inefficacités de production (rendues plus ou moins possibles selon les largesses obtenues lors de la négociation budgétaire).

Pour évaluer l'effet potentiel d'un système de tarification par pathologie sur la dépense hospitalière, nous adoptons une démarche empirique en deux étapes.

– La première étape consiste à effectuer une analyse économétrique des coûts hospitaliers de façon à repérer, dans la variabilité des coûts, ce qui peut être attribué à des inefficacités. La méthode retenue correspond dans son principe aux approches en termes de frontières de production : des différences de coût, repérées toutes choses égales par ailleurs, entre deux établissements, sont identifiées comme reflétant différents niveaux d'efficacité.

– Dans la seconde étape on évalue, sur l'échantillon utilisé, le niveau de budget qui aurait suffi dans le cas où les coûts auraient correspondu à une activité efficace. Sous l'hypothèse qu'une tarification par pathologie permettrait d'obtenir ces niveaux d'efficacité, les résultats obtenus s'interprètent comme les effets simulés des gains budgétaires potentiels liés à l'instauration d'un tel système de tarification. De plus, les coûts de traitement efficaces estimés permettent de définir des formules de paiement que nous préconisons.

Analyse de la variabilité des coûts hospitaliers

Des données à trois dimensions : séjour-hôpital-année

Les données que nous utilisons sont des données individuelles de séjours effectués pour infarctus du myocarde aigu dans les hôpitaux publics français. On dispose d'un échantillon de 7 314 séjours observés dans 36 hôpitaux sur les années 1994-

1997. Ces données sont issues de la base de coût constituée par le Programme de médicalisation des systèmes d'informations (PMSI). Dans le PMSI, les séjours sont classés en Groupes homogènes de malades (GHM) à l'aide d'une arborescence sur la base des différents diagnostics et actes pratiqués sur le patient. Un GHM définit la pathologie pour laquelle la tutelle est supposée fixer une tarification. Afin que les séjours observés soient les plus homogènes possible en termes de pathologies, nous avons sélectionné les séjours de patients âgés d'au moins 40 ans, pour lesquels le diagnostic principal est l'infarctus du myocarde aigu et qui ont été classés dans le même GHM : infarctus du myocarde sans complication (GHM 179).

Pour chaque séjour, on observe le coût du séjour, les diagnostics secondaires, les actes pratiqués, le mode d'entrée (en provenance du domicile ou d'un autre hôpital), le mode de sortie (sortie vers le domicile, un autre hôpital ou un autre service), la durée du séjour, l'âge et le sexe du patient.

Les données du PMSI permettent d'accéder à une information d'une richesse considérable : les observations, effectuées au niveau du séjour, sont très détaillées en ce qui concerne les diagnostics, les traitements et les coûts associés. Le PMSI offre toutefois une vision restrictive de la production hospitalière et l'information sur la qualité des soins reste pauvre. Les seules observations sur l'output sont constituées des modalités de la variable « mode de sortie », c'est-à-dire le décès, le transfert ou le retour au domicile. On ne dispose pas d'informations relatives à la qualité de vie du malade après le séjour, à une réhospitalisation précoce, ni à la contraction d'une éventuelle infection nosocomiale. Enfin, on ne dispose pas de renseignement sur la qualité du service rendu en termes d'accueil, de confort et de prise en charge de la douleur.

Les données de la base nationale PMSI couvrent exhaustivement les hôpitaux publics français. En revanche, les données de la base de coût sont particulières dans la mesure où la participation à sa constitution est définie sur le principe du volontariat. Les établissements, en nombre limité, qui ont accepté de fournir les informations sur leurs coûts par séjour

disposent par définition d'une bonne comptabilité analytique. Toutefois, des analyses comparatives avec la base exhaustive¹ ont permis de vérifier que la base de coût est représentative - pour la pathologie étudiée et les actes pratiqués - des séjours effectués dans les hôpitaux publics français (Dormont et Milcent, [2002b]).

La structure des données est assez complexe. Tout d'abord, il s'agit de données stratifiées à trois dimensions : séjour-hôpital-année. Ensuite, les données sont non cylindrées dans plusieurs dimensions : non seulement le nombre de séjours observés varie selon l'hôpital et l'année considérés, mais la période d'observation des hôpitaux est de longueur variable.

Spécification économétrique de la fonction de coût

Afin de limiter la sélection des patients, nous défendons le principe d'un affinement de la tarification par l'intégration de caractéristiques individuelles observables associées au séjour. Par ailleurs, la tutelle attribuée aux établissements des missions différentes et doit intégrer dans le paiement le fait d'assumer des tâches d'enseignement et de recherche, d'être un hôpital de proximité, etc. Grâce à nos données à trois dimensions, nous pouvons spécifier un modèle permettant d'intégrer dans la tarification les caractéristiques observables des patients et les hétérogénéités entre les hôpitaux. En outre, il est possible d'identifier finement une composante d'aléa moral transitoire et de prendre en compte un paramètre d'hétérogénéité non observée entre les établissements pouvant refléter (entre autres) la qualité des soins.

Soit $C_{i,b,t}$ le coût du séjour i effectué dans l'hôpital b pendant l'année t . Nous avons retenu pour la fonction de coût la spécification suivante :

$$C_{i,b,t} = X'_{i,b,t}\gamma_t + W'_{b,t}\alpha + Q'_b\lambda + a + c_t + \eta_b - \varepsilon_{b,t} + u_{i,b,t} \quad (1)$$

$X'_{i,b,t}$ représente les caractéristiques individuelles observables des patients : effets croisés âge-sexe, mode d'entrée, mode de sortie, durée de séjour. Les variables $W'_{b,t}$ et Q'_b sont les caractéristiques observables de l'hôpital, constantes ou variables dans le temps : catégorie de l'établissement, technicité, taux d'entrée et de sortie par transfert, taux d'application des actes innovants. a est une constante.

A caractéristiques des patients données, la variabilité des coûts peut trouver sa source dans les caractéristiques des hôpitaux : existence d'économies d'échelle ou de gamme, missions qui peuvent être assurées (CHR, hôpital de proximité), diversification des pathologies traitées, qualité des soins délivrés (mise en œuvre des technologies innovantes, accueil, prise en charge de la douleur, etc.), niveau de capital humain du personnel soignant, qualité de la gestion des ressources humaines, rigueur dans la gestion de l'hôpital. Certains de ces déterminants sont observables, d'autres non.

Dans notre approche, nous supposons que la tutelle dispose des données du PMSI pour établir les paiements hospitaliers : le partage entre variables observables et non observables est alors le même pour la tutelle et l'économètre. Les caractéristiques observables des patients sont les variables $X'_{i,b,t}$ et celles des hôpitaux les variables $W'_{b,t}$ et Q'_b . A caractéristiques observables données, la variabilité des coûts dépend, dans la spécification (1), du terme

$$c_t + \eta_b - \varepsilon_{b,t} + u_{i,b,t}$$

c_t est un effet fixe temporel, commun à tous les établissements, qui est lié à la progression générale des budgets hospitaliers et reflète l'évolution des prix, des salaires et du progrès technique. Dans le contexte institutionnel français en vigueur depuis les ordonnances de 1996, c_t dépend aussi de la progression des budgets hospitaliers qui découle du vote de l'ONDAM (Objectif national de dépenses en assurance maladie), dans le cadre de la loi de financement de la Sécurité sociale.

L'hétérogénéité non observée des patients est spécifiée par la perturbation $u_{i,b,t}$. L'interprétation de la perturbation $\varepsilon_{b,t}$ est étudiée ci-dessous.

Interprétation de l'effet spécifique hôpital η_b

L'hétérogénéité non observée des établissements est spécifiée comme un effet spécifique hôpital, qui peut être fixe ou aléatoire. Elle peut être conçue comme la résultante de trois composantes :

$$\eta_b = \eta_b^{as} + \eta_b^{hm} + \eta_b^q$$

Dans le cadre théorique d'une relation d'agence entre la tutelle et l'hôpital, où la tutelle observe mal l'effort de réduction du coût fourni par le gestionnaire de l'hôpital (aléa moral) et les caractéristiques de l'établissement expliquant sa productivité (antisélection), les composantes de η_b peuvent s'interpréter de la façon suivante. η_b^{as} est un paramètre d'antisélection : les infrastructures de l'hôpital conduisent à un fonctionnement plus ou moins coûteux, il peut exister des économies d'échelle ou de gamme². η_b^{hm} représente le hasard moral de long terme : la gestion de l'hôpital peut être dispendieuse de façon permanente. Enfin, η_b^q correspond à la qualité : performances des soins, accueil, prise en charge de la douleur, etc.

Interprétation du terme $\varepsilon_{b,t}$

La variabilité des effets spécifiques η_b reflète les écarts de coûts moyens entre les établissements, à caractéristiques observables $X'_{i,b,t}$, $W'_{b,t}$ et Q'_b données. Comment maintenant interpréter la perturbation $\varepsilon_{b,t}$? À caractéristiques observables et à choc conjoncturel (c_t) donnés, elle est définie comme l'écart, une année t , de l'hôpital b à son niveau de coût moyen. En tant que tel, $\varepsilon_{b,t}$ peut donc être compris comme un indicateur de l'effort transitoire de réduction du coût, autrement dit, de l'aléa moral correspondant à l'effort fourni par le gestionnaire de l'hôpital en matière de réduction des coûts. Concrètement,

on peut penser à la plus ou moins grande rigueur qui peut être adoptée par celui-ci dans les procédures de négociation pour les marchés des consommables et pour les tarifs des différentes activités assurées par des intervenants extérieurs. Pour se référer à nouveau au contexte institutionnel du budget global, $\varepsilon_{h,t}$ correspond à l'inefficacité « autorisée », compte tenu de la contrainte budgétaire plus ou moins lâche subie par l'hôpital h en t .

En principe, $\varepsilon_{h,t}$ contient aussi les composantes ordinaires de toute perturbation : variables omises et erreurs de mesure. Ces dernières doivent toutefois être d'une importance modérée.

Par définition, $\varepsilon_{h,t}$ est en effet la perturbation d'une équation qui explique le coût de tous les séjours : dans ce cadre, une erreur de mesure figurant dans $\varepsilon_{h,t}$ devrait affecter systématiquement tous les enregistrements des séjours pour l'hôpital h à l'année t . Il ne peut donc s'agir d'une erreur de retranscription pour un patient, mais d'un biais systématique, une année donnée, dans les critères utilisés pour définir la durée d'un séjour par exemple, ou encore d'une erreur sur la catégorie d'établissement à laquelle appartient l'hôpital. Cette dernière hypothèse est peu vraisemblable.

Examinons maintenant l'hypothèse des variables omises qui pourraient figurer dans $\varepsilon_{h,t}$: elles correspondent à des chocs ayant affecté l'hôpital h et lui seul, une année t . Il peut s'agir d'une panne touchant son système d'eau courante (pour prendre un exemple parisien fameux). Nous pensons que la tutelle a intérêt à assimiler *a priori* ces accidents à du hasard moral, afin de pousser les établissements à les déclarer, dans le cas où ils estimeraient que les écarts de coûts auxquels ils conduisent sont exceptionnels et justifiables.

On peut donc interpréter la perturbation $\varepsilon_{h,t}$ comme un indicateur de l'aléa moral transitoire, lié aux efforts fournis par le gestionnaire de l'hôpital en matière de réduction des coûts.

Nous avons vu que les caractéristiques des hôpitaux η_h comportent aussi, vraisemblablement, de l'aléa moral de long terme. De ce fait, notre estimation ne permettra pas de repé-

rer « tout » l'aléa moral en estimant la variance de $\varepsilon_{h,t}$, mais seulement l'aléa moral « transitoire ». En tout état de cause, les développements qui précèdent permettent raisonnablement de penser que la variance de $\varepsilon_{h,t}$ est exclusivement due à de l'aléa moral. Puisque nous pouvons identifier cette composante, il sera possible de définir une formule de paiement permettant de l'éliminer.

Influences des caractéristiques observables, de l'aléa moral et de l'hétérogénéité entre les établissements

La fonction de coût a été estimée de façon différente selon que des effets spécifiques hôpitaux η_h étaient supposés fixes ou aléatoires. Un test nous conduit à privilégier le modèle à effets fixes (Dormont et Milcent, [2002_a]) mais il est intéressant de commenter les résultats obtenus sur les deux spécifications.

Les coefficients estimés pour les caractéristiques individuelles confirment des résultats bien connus en matière de pathologie cardiaque : le coût diminue avec l'âge ; il est de façon générale plus élevé pour les hommes. Une journée supplémentaire d'hospitalisation coûte en moyenne, toutes choses égales par ailleurs, 380 € environ. Par ailleurs, l'estimation d'une spécification incomplète n'incluant que les caractéristiques individuelles des patients comme variables explicatives permet de constater que 54,2 % de la variance des coûts des séjours peut être expliquée par l'hétérogénéité observable des patients.

Les coefficients estimés des caractéristiques observables des hôpitaux montrent, lors de l'estimation du modèle à effets aléatoires, que les coûts des centres hospitaliers régionaux (CHR) et des hôpitaux privés PSPH (participant au service public hospitalier) ne sont pas significativement différents de ceux des autres hôpitaux publics. Ce résultat est surprenant. Concernant les CHR, les experts de la mission PMSI évaluent un surcoût de 13 % lié aux activités d'enseignement et de recherche dans les hôpitaux français (direction des hôpitaux,

[1996]). Sur un échantillon d'hôpitaux espagnols, Lopez-Casasnovas et Saez [1999] estiment un surcoût significatif, de 9 % environ. Quant aux hôpitaux privés PSPH, une étude de la Fédération des établissements hospitaliers et associations privées à but non lucratif déclare un écart de 14 % sur les charges salariales des PSPH qui se traduirait par un alourdissement de 7 % de leurs budgets (Apparitio, Brocas et Moisson, [1999]).

En revanche, on obtient un coefficient positif et significatif (431 €) pour la variable TECH, variable indicatrice du fait que l'hôpital a effectué les investissements nécessaires pour la réalisation d'actes innovants. Ce coefficient peut s'interpréter en calculant le surcoût relatif correspondant, par rapport au coût moyen du séjour de référence (4 198 €). Les coûts de ces hôpitaux techniques sont donc supérieurs de 10,2 %. Cet effet positif est observé selon des modalités différentes en fonction de la spécification considérée. Au total, être un hôpital technique conduit toujours à un surcoût. Le résultat concernant l'absence d'impact significatif des activités d'enseignement et de recherche doit alors être interprété au regard de l'effet positif de cette variable TECH. En effet, tous les CHR de notre échantillon sont des hôpitaux techniques sur toute leur période d'observation. Le résultat obtenu signifie donc que des hôpitaux qui n'ont pas le statut de CHR mais pratiquent les activités innovantes ont aussi des coûts plus élevés : l'existence d'un surcoût est plus directement liée à la réalité d'une activité de pointe (TECH) qu'au statut.

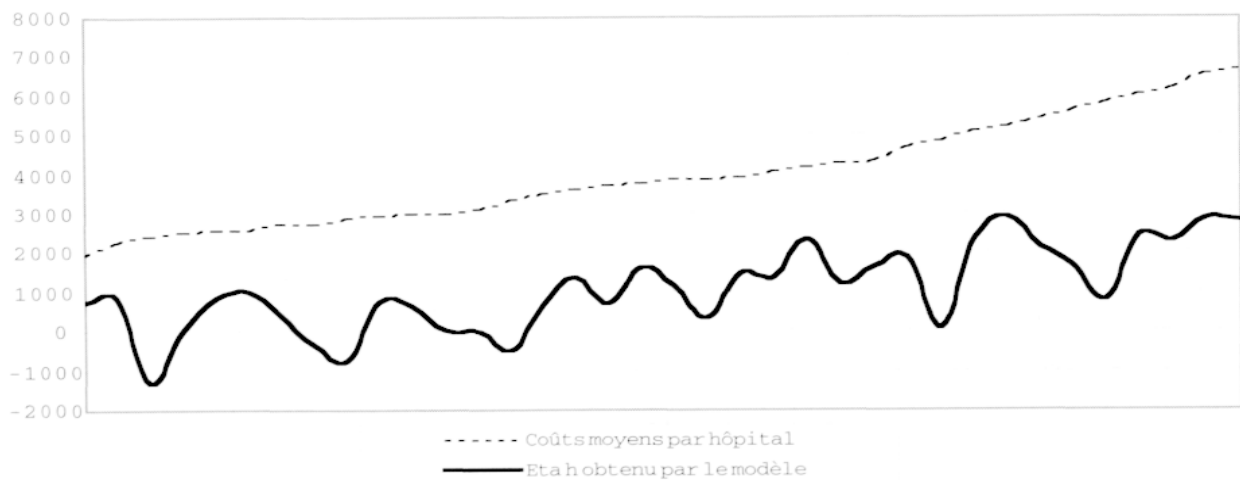
Lorsque l'on considère le modèle à effets fixes, la variable TECH n'est plus significative : les effets fixes η_h prennent en compte les différences permanentes dans les coûts moyens par hôpital.

Compte tenu des tests réalisés, la spécification avec des effets fixes hôpitaux permet d'obtenir des estimations convergentes des termes η_h et $\varepsilon_{h,t}$ et de leurs écarts-types respectifs σ_η et σ_ε . On constate que la corrélation entre les effets estimés $\hat{\eta}_h$ et $\hat{\varepsilon}_{h,t}$ est très faible et non significative. Pour un des modèles

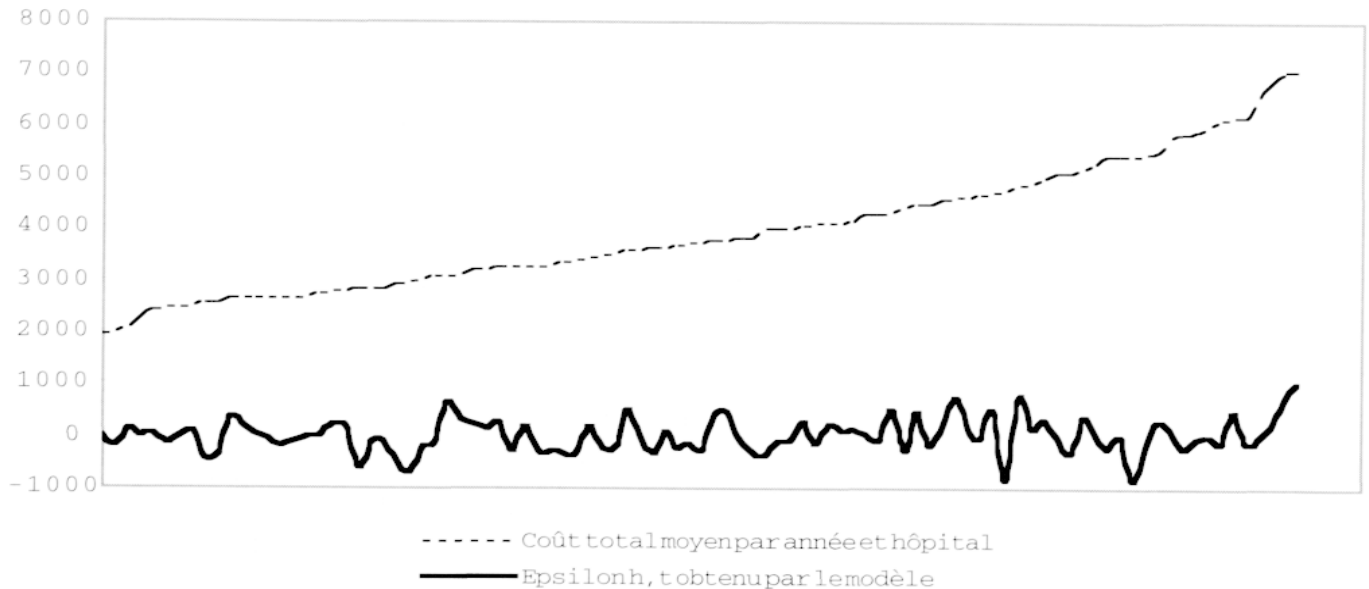
considérés, on trouve $\hat{\sigma}_\varepsilon = 399 \text{ €}$ et $\hat{\sigma}_\eta = 1\,057 \text{ €}$. L'importance de ces écarts-type peut être évaluée en les rapportant à l'écart-type des coûts de nos séjours : 2 863 €, pour un coût moyen de 4198 €. Bien que nous en ayons retenu une définition très restrictive, l'importance de la variabilité des coûts due au hasard moral transitoire est ainsi considérable : σ_ε représente environ 40 % de σ_η . Ce résultat suggère que la mise en place d'une tarification incitant à la réduction de cet aléa moral transitoire devrait conduire à de fortes économies budgétaires.

Les graphiques n° 1 et 2, mettent en rapport les effets estimés $\hat{\eta}_h$ et $\hat{\varepsilon}_{h,t}$ avec les coûts moyens par hôpital ou par hôpital-année correspondants. Les observations sont rangées par ordre croissant de coûts moyens. On constate que les effets spécifiques hôpitaux sont en rapport avec la hiérarchie des coûts moyens, mais sont très loin de l'expliquer entièrement (graphique n° 1). Le graphique n° 2 montre bien la régularité des coûts, face aux fluctuations de l'aléa moral transitoire : face à une réalité fluctuante de cas traités, plus ou moins coûteux en moyenne par hôpital-année, la régularité des budgets autorise de nombreuses inefficacités.

Graphique 1



Graphique 2



Economies budgétaires potentielles

Les estimations permettent de simuler l'application des deux formules de tarifications P^1 et P^2 que nous proposons (les $\hat{\cdot}$ désignent les estimations obtenues) :

$$P^1_{i,b,t} = X'_{i,b,t} \hat{\gamma}_t + W'_{h,t} \hat{\alpha} + \hat{a} + \hat{c}_t + \hat{\eta}_b + \widehat{Min}_{h,t} (-\varepsilon_{b,t})$$

$$P^2_{i,b,t} = X'_{i,b,t} \hat{\gamma}_t + W'_{h,t} \hat{\alpha} + \hat{a} + \hat{c}_t + \widehat{Min}_{h,t} \{\eta_b - \varepsilon_{b,t}\}$$

Les deux formules de paiement prennent en compte les caractéristiques observables des patients et des établissements.

– La tarification P^2 est basée sur l'hypothèse que toute hétérogénéité non observée entre les établissements correspond à de l'aléa moral : en se calant sur l'hôpital qui a la valeur minimale de $\eta_b - \varepsilon_{b,t}$, P^2 écrase toutes les hétérogénéités non observées $\eta_b - \varepsilon_{b,t}$, alors que η_b peut correspondre pour partie à de bonnes caractéristiques, une certaine qualité des soins par exemple.

– La tarification P^1 est plus souple. Avec P^1 , la tutelle tient compte des hétérogénéités non observées constantes dans le temps η_h , qu'elles soient dues à une mauvaise gestion ou à une qualité particulière des soins. Elle rémunère toutes les hétérogénéités entre les établissements : les différences de qualité mais aussi les écarts de coûts qui seraient dus à des inefficacités permanentes de gestion. Cette tarification reste incitative car elle ne rémunère pas les variations de coûts moyens hospitaliers dus à l'aléa moral purement transitoire $\varepsilon_{h,t}$.

D'après nos simulations, la tarification P^1 conduit à des économies potentielles de 16 % environ et la tarification P^2 à des gains budgétaires de 42 à 46 %.

La stratégie de tarification (P^1) doit être recommandée. Prudente, elle permet d'éviter de niveler les paiements sur un établissement dont la qualité des prestations (variable difficilement vérifiable par la tutelle) serait la plus mauvaise. Elle conduit à un résultat non négligeable en matière d'économie budgétaire (16 %) car elle est suffisamment incitative pour réduire une grande part de l'aléa moral.

On a vu que l'hétérogénéité des patients expliquait plus de 50 % de la variance des coûts. Ne pas la prendre en compte dans les formules de tarification serait une incitation à la sélection des patients par les hôpitaux. Les formules proposées limitent ce risque grâce à une prise en compte des caractéristiques individuelles des patients. Cette prise en compte correspond à un paiement prospectif à travers une grille de tarification de la forme : « pour un homme, dans telle tranche d'âge, vous recevrez X € en supplément ». Dans cette perspective, la formule P^1 peut être considérée comme un paiement prospectif, amendé par la forme d'ajustement au risque que constitue la prise en compte des hétérogénéités observées des patients. En plus, la formule P^1 comporte une part de paiement rétrospectif dans la mesure où l'on rémunère les écarts de coûts dus à η_h .

On peut, selon ces critères, distinguer dans la formule de P^1 les éléments de paiements prospectifs et rétrospectifs. Considérons l'expression classique des paiements mixtes comme

une moyenne pondérée du forfait F et du remboursement au coût observé C : $P = \mu F + (1-\mu)C$. A partir de cette décomposition et de notre simulation des paiements P^i , on peut évaluer les poids μ . On obtient au niveau de l'échantillon une moyenne $\hat{\mu} = 44,7 \%$, avec un écart-type de $12,8 \%$. Soulignons que cette évaluation n'est en aucune manière une formule de paiement. Elle est effectuée *a posteriori* à partir des simulations de paiements réalisées sur l'échantillon et permet de connaître, à titre indicatif seulement, l'ordre de grandeur de la prise en compte du rétrospectif que représenterait un paiement défini selon la formule P^i , qui incorpore une rémunération des écarts η_{ij} .

Nous avons interprété l'aléa moral transitoire $\varepsilon_{h,t}$ comme reflétant l'effort fourni par le gestionnaire de l'hôpital en matière de réduction des coûts en se référant à la plus ou moins grande rigueur dans la négociation des marchés pour les consommables et les prix des activités externalisées. L'ordre de grandeur de l'économie budgétaire simulée, 16% , peut-il être considéré comme excessif, au regard de cette interprétation de $\varepsilon_{h,t}$? En aucune manière : le prix des consommables est un élément très important des coûts de traitement. Pour prendre un exemple, la réalisation d'un acte innovant comme l'angioplastie entraîne un surcoût de 60% , surcoût attribuable pour l'essentiel au matériel utilisé.

Intérêt et limite des tarifications proposées

Les inconvénients d'un système de tarification hospitalier purement prospectif sont connus : sélection des patients, baisse de la qualité des soins. Pour les éviter, de nombreux auteurs ont préconisé l'utilisation d'une tarification mixte, qui combinerait forfait et remboursement du coût observé. Notre analyse économétrique permet de définir une tarification qui tient compte de l'hétérogénéité des hôpitaux.

Celle-ci est prise en compte par des caractéristiques observables des établissements et des effets spécifiques hôpitaux. Ces derniers peuvent être interprétés comme les résultantes de trois effets : la sélection adverse, le hasard moral de long terme et les différences permanentes dans la qualité des soins. On propose deux formules de paiement. La première prend en compte toute l'hétérogénéité non observée entre les hôpitaux, pourvu qu'elle soit constante dans le temps. La seconde ignore toute hétérogénéité non observée.

La première formule de paiement semble préférable : elle présente le grand avantage de procurer un surcroît de financement aux hôpitaux délivrant des soins de haute qualité. Par ailleurs, nous montrons qu'elle est susceptible de conduire à des économies budgétaires substantielles, car elle est suffisamment incitative pour réduire la partie – non négligeable – des coûts attribuables au hasard moral transitoire. Ces économies sont obtenues par une pression ne jouant que sur le gestionnaire de l'hôpital. En principe, le gain est acquis sans contrainte directe sur les patients ni sur le personnel soignant.

Cette tarification est facile à mettre en œuvre, dès lors que la tutelle a accès à l'information sur les coûts hospitaliers, établissement par établissement. A cet égard les économies budgétaires que nos simulations mettent en évidence doivent être mises en regard des coûts de gestion d'un système de tarification à la pathologie, notamment le coût du système d'information qu'il nécessite. Mais il faut bien comprendre qu'en France, ce système d'information est déjà installé (le PMSI) et qu'il n'y aurait qu'un coût marginal à le compléter pour qu'il contienne des informations sur les coûts par séjours pour tous les hôpitaux.

Un inconvénient de ce système est qu'il conduit à attribuer des financements plus élevés à des hôpitaux qui sont plus coûteux de façon permanente à cause d'une mauvaise gestion. Notre méthode permet d'identifier la composante de la variabilité des coûts due au hasard moral transitoire. Mais nous ne pouvons pas discriminer entre les composantes désirables et indésirables de l'hétérogénéité non observée. De ce fait, la tutelle doit choisir entre un paiement qui ignore toute hétérogénéité non observée et un paiement qui finance toute l'hé-

térogénéité non observée. Le choix entre ces deux paiements dépend des poids accordés à l'efficacité et à la qualité des soins dans la fonction d'utilité sociale.

L'intérêt de notre approche réside dans le fait qu'elle exploite l'information disponible au niveau individuel des séjours et pour une pathologie bien circonscrite. Se concentrer sur une affection comme l'infarctus du myocarde aigu permet d'étudier les composantes de la variabilité des coûts hospitaliers, mais aussi de mettre en lumière le rôle des actes innovants en matière de coût et de qualité des soins et les enjeux et difficultés liés à l'intégration des innovations dans une grille de tarification par pathologie (*cf.* Milcent, [2002]). Il est clair que le mode de régulation auquel les établissements sont soumis influence les taux d'application et le rythme de diffusion des innovations techniques (Delattre et *alii*, [2002]). Concernant les innovations, une tarification trop homogène peut conduire à des comportements de type sélection ou discrimination de patients. En effet, nos estimations montrent que les augmentations de coûts qui résultent de la pratique des actes innovants sont assez considérables : de 34 % à 61 % ! De telles évaluations impliquent que l'on ne peut pas introduire une tarification par pathologie sans prendre les précautions spécifiques visant à éviter de pénaliser les hôpitaux pratiquant des actes qui sont considérés comme des progrès en matière de diagnostic et de soins.

Ce travail comporte certaines limites, qui tiennent tout d'abord à la qualité des données : la base de coût PMSI ne contient qu'un nombre restreint d'hôpitaux, tous volontaires ; nous n'observons pas le devenir du patient au-delà de son séjour ; les informations sur la qualité des services rendus sont quasi inexistantes. Nous avons pu remédier en partie au manque d'information sur la qualité des soins grâce au fait que nos données de panel stratifiées permettent d'évaluer un effet fixe hôpital, vraisemblablement en rapport avec la qualité. Toujours est-il que la collecte de l'information sur les coûts des séjours hospitaliers devrait être effectuée sur un champ beaucoup plus large et représentatif des établissements. Il conviendrait en outre, au niveau des données PMSI, d'améliorer l'information sur la qualité des soins et le suivi du patient.

Nos exercices de simulation comportent aussi leurs limites. Ils sont effectués en supposant que les coûts hospitaliers correspondent effectivement aux niveaux de coûts efficaces définis par la tarification. Leur portée pratique dépend donc de la crédibilité d'une menace de fermeture de l'établissement en cas de déficits répétés, mais aussi des marges de manœuvre dont disposent les établissements pour s'adapter, en l'absence d'autonomie dans la gestion de leur budget et de leur personnel. Par ailleurs, nos simulations sont effectuées à comportement et à niveau d'activité constants : trouver 16 % d'économies signifie que, pour l'activité observée ces années-là, on aurait pu dépenser 16 % de moins. Or, à la différence d'une régulation par budget global, le volume d'activité n'est en principe pas borné dans la tarification par pathologie. Cette caractéristique est d'ailleurs à mettre au crédit de ce mode de tarification : il devrait inciter les hôpitaux à un fonctionnement efficace sans induire une pénurie de moyens. Il devrait être considéré comme un progrès par les acteurs du système hospitalier, car il permet de répondre aux besoins de la population en instaurant une contrainte budgétaire beaucoup moins brutale que la dotation globale. En contrepartie, la maîtrise par la tutelle de la dépense hospitalière globale serait moins assurée, ce qui renforce l'impératif de la définition rigoureuse d'une tarification incitant les établissements à fonctionner de façon efficace.

Cet article est issu d'une étude financée en partie par une convention avec la DREES (Ministère de l'emploi et de la solidarité). Nous remercions Werner Antweiler (University of British Columbia) pour ses précieux commentaires, ainsi que Bruno Crépon, Alberto Holly et Nathalie Destais. Nous avons aussi profité des encouragements de Pierre Morin et des remarques des participants du séminaire santé CREST-IEI, du séminaire Théma et du séminaire Fourgeaud.

Brigitte Dormont est professeur de sciences économiques à l'université Paris X-Nanterre et est membre du Théma (CNRS, UMR 7536).

Adresse : Université Paris X-Nanterre, UFR SEGMI, bâtiment G, 200 avenue de la République 92001 Nanterre cedex. E-mail : dormont@u-paris10.fr. Tél : 01 40 97 78 36.

Carine Milcent est chargé de recherche au CNRS et est membre du Delta (CNRS, UMR 8545).

Adresse : DELTA-ENS 48, 48, bd Jourdan, 75014 Paris. E-mail : milcent@delta-ens.fr. Tél : 06 03 99 58 06.

Notes

1. Pour laquelle les coûts ne sont pas observés.

2. En utilisant des données d'enquête, nous avons essayé de saisir ces effets directement en intégrant dans le modèle

la taille de l'établissement et la diversification de son activité, sans obtenir de résultat significatif, sans doute à cause du fait que l'on étudie ici le coût des séjours pour une seule pathologie.

Références

- S. Apparitio, A.-M. Brocas, et J.-C. Moïsdon [1999] : *La place du PSI dans l'allocation des ressources en Île-de-France*, Agence régionale d'hospitalisation d'Île-de-France - Rapport technique.
- E. Auriol et J.J. Laffont [1992] : *Regulation by Duopoly*, **Journal of Economics and Management Strategy**, vol 1, n° 3, pp 507-533.
- E. Delattre, B. Dormont, M. McClellan, C. Milcent [2002] : *Systèmes de tarification et évolutions de la variabilité des coûts hospitaliers en France et aux Etats-Unis*. La tarification à la pathologie : les leçons de l'expérience étrangère. Dossiers Solidarité et Santé, hors série. La Documentation française, juillet.
- Direction des hôpitaux de Paris, mission PMSI [1996] : *Le PMSI, analyse médico-économique de l'activité hospitalière*, La lettre d'information hospitalières, numéro spécial.
- B. Dormont et C. Milcent [2002a] : *Effets potentiels d'une tarification par pathologie : analyse économétrique des coûts hospitaliers*, doc. de travail Théma, mai.
- B. Dormont et C. Milcent [2002b] : *Quelle régulation pour les hôpitaux publics français ?*, Rapport pour la DREES, ministère de l'Emploi et de la solidarité, convention d'étude n° 99130201.
- R. P. Ellis [1998] : *Creaming, Dumping, Skimming: Provider Competition on the Intensive and Extensive Margins*, **Journal of Health Economics**, vol 17 : pp 537-555.
- L. Hartmann, K. Moschetti et L. Rochaix-Ranson [2001] : *La tarification à la pathologie aux Etats-Unis*, **Solidarité et santé**, n° 1, pp 17-31.
- E. B. Keeler [1990] : *What Proportion of Hospital Cost Differences is Justifiable ?*, **Journal of Health Economics** 9(3), pp. 359-365.
- J. J. Laffont et J. Tirole [1993] : **A Theory of Incentives in Procurement and Regulation**, MIT Press.
- G. Lopez-Casasnovas et M. Saez [1999] : *The Impact of Teaching Status on Average Costs in Spanish Hospitals*, **Health Economics**, vol 8, n° 7: pp 641-651.
- A. C. T. Ma [1994] : *Health Care Payment Systems: Cost and Quality Incentives*, **Journal of Economics and Management Strategy**, vol 3, n° 1: pp 93-112.
- A. C. T. Ma [1998] : *Health Care Payment Systems: Cost and Quality Incentives-Reply*, **Journal of Economics and Management Strategy**, vol 7, n° 1 : pp 139-142.
- M. McClellan [1997] : *Hospital Reimbursement Incentives : an Empirical Analysis*, **Journal of Economics and Management Strategy**, 6(1) : pp. 91-128.
- C. Milcent [2002] : *Impact des innovations sur la durée et le coût des séjours hospitaliers*, in Dormont et Milcent [2002b], chap. 3.
- J.C Moïsdon et D. Tonneau [1997] : *Gestion externe et gestion interne du système hospitalier français. Tendances actuelles*, Dix ans d'avancées en économie de la santé, Actes des XIX^e Journées des économistes de la santé français. Ed : John Libbey.
- M. Mougeot [1999] : *Régulation du système de santé*, Conseil d'analyse écono-

mique. La Documentation française, Paris.

J.P. Newhouse [1996] : *Reimbursing Health Plans and Health Providers : Efficiency in Production versus Selection*. **Journal of Economic Literature**, vol. XXXIV : pp. 1236-1263.

G. Pope [1990] : *Using Hospital-Specific Costs to Improve the Fairness of Prospective Reimbursement*. **Journal of Health Economics**, vol 9, n° 3 : pp 237-251.

A. Shleifer [1985] : *A Theory of Yardstick Competition*, **Rand Journal of Economics**, vol 16 : pp 319-327.