

Biais d'endogénéité et tarification à la pathologie

Romain Lesur

CREST-LEI, THEMA and ENSAE *

Version très provisoire. Ne pas citer.

22 septembre 2003

1 Introduction

Les réformes de la tarification hospitalière menées dans les principaux pays développés lors de ces vingt dernières années ont donné lieu à une vaste littérature théorique et empirique afin d'évaluer les incitations des différents modes de paiement. Au début des années quatre-vingts, la plupart des pays ont abandonné les modes de financement assis sur une logique de remboursement des coûts, jugés trop inflationnistes. La caractéristique commune à ces deux modes de financement qui ont alors été mis en place par les différents régulateurs ont été le budget global et la tarification à l'activité. Ces deux types de financement revêtent un caractère prospectif : les hôpitaux savent par avance que les montants versés seront indépendants des coûts de traitement finaux. Ces financements prospectifs ont par essence des vertus incitatives. Néanmoins, le budget global et la tarification à l'activité connaissent chacun leurs limites.

Avec le budget global, le régulateur fixe par avance le budget de fonctionnement de l'hôpital. Le problème du budget global est le manque d'engagement du financeur à faire évoluer le financement de l'hôpital en fonction de son activité. Avec le temps, ce type de régulation mène à des disparités budgétaires très fortes entre établissements, sources d'iniquité, mais peut aussi inciter les hôpitaux à sélectionner les patients sur la base d'informations médicales les concernant.

*Address : Romain Lesur - ENSAE - 3, av Pierre Larousse - 92245 Malakoff Cedex - France - phone : (+33)141 17 65 29 - fax : (+33)141 17 38 52 - email : lesur@ensae.fr

En théorie, la tarification à l'activité ne présente pas l'écueil précédent car le montant du paiement prospectif est conditionnel à l'état de santé du patient considéré : elle est fondée sur une classification *médicale* des séjours, dont la version française se nomme la classification en Groupes Homogènes de Malades (GHM). Un tarif est calculé pour chaque classe de séjours, et l'hôpital est rémunéré en fonction de ces "prix". Un des aspects importants de la tarification à l'activité concerne la fixation du tarif de chaque GHM : fixer ce montant au niveau du coût espéré permet d'annuler les rentes laissées aux hôpitaux et alors d'implémenter l'optimum. Un consensus s'est donc établi concernant la détermination du "prix" par GHM : il convient de le fixer au niveau coût de traitement espéré. D'un point de vue méthodologique, on peut estimer le coût de traitement espéré en calculant le coût moyen ou le coût médian¹ par GHM.

Toutefois, on constate empiriquement que l'état de santé du patient à l'entrée de l'hôpital ne détermine pas complètement le coût final. D'autres facteurs, comme le choix de traitement conditionne fortement le coût de traitement final. Cela a mené le régulateur à créer des classes de séjours dépendant du traitement pratiqué : pour les accouchements, il existe des GHM "Accouchements par voie basse" (GHM 533, 539 et 540) et "Accouchements par césariennes" (GHM 530 et 531). McClellan (1997) a montré que l'utilisation du traitement (variable de nature rétrospective) dans la formule de paiement réduit le caractère prospectif de la tarification à l'activité. Sur ce point précis, la tarification à l'activité fait donc moins bien que le budget global, dans lequel le producteur internalise complètement les coûts liés au choix de traitement.

Par ailleurs, on attend sur ce point que les gains d'efficacité procurés par la tarification à l'activité la rende préférable au remboursement des coûts. Nous montrons dans ce papier que la stratégie usuelle de fixation des "prix" décrite précédemment peut induire un recours aux traitements lourds plus fréquents que lorsque les hôpitaux sont remboursés au niveau de leurs coûts. Cette "surutilisation" des traitements lourds représente alors une inefficacité qui relativise le gain social de la tarification à l'activité par rapport au remboursement des coûts.

L'origine de cette inefficacité réside dans la corrélation positive entre le coût de traitement final et la gravité de la pathologie : ce biais d'endogénéité implique une surestimation du coût d'opportunité des traitements lourds, et incite les hôpitaux à y recourir plus fréquemment.

Le problème que nous soulevons s'ajoute à la liste des limites connues

¹si la distribution des coûts est symétrique, la médiane et la moyenne sont toutes deux des estimateurs de la tendance centrale.

de la tarification à l'activité, qui sont la sélection des patients, le surcodage et le transfert des patients. Le problème de la sélection n'est pas totalement résolu s'il subsiste une asymétrie d'information en faveur de l'hôpital sur l'état de santé du patient (Ma (1994) et Chalkley et Malcomson(2002)). On peut réduire les incitations à la sélection en légiférant, en augmentant le tarif de chaque GHM ou en pratiquant le partage des coûts.

Le surcodage provient de la nature même de l'information médicale qui n'est pas toujours vérifiable par le régulateur : Silverman et Skinner (2001) étudient les cas de surcodage des "pneumonies" vers les "infections respiratoires". Ces deux pathologies sont proches et délicates à distinguer, et la seconde est rémunérée 2000 USD de plus que la première. D'un point de vue théorique, ce phénomène lié à l'asymétrie et l'hétérogénéité de l'information est classique et l'arbitrage en jeu se situe entre extraction des rentes informationnelles et efficacité.

Le transfert précoce des patients provient directement de l'"unité" utilisée par le régulateur pour rémunérer les hôpitaux, à savoir le séjour. Réduire la durée de séjour est donc le moyen le plus aisé pour réduire ses coûts. Par conséquent, les hôpitaux sont incités à transférer vers un autre établissement, qui recevra également une rémunération pour ce même patient. La solution adoptée par Medicare est de rémunérer suivant la durée du séjour les patients qui auront fait l'objet d'un transfert (Chalkley et Malcomson (2002)).

La suite de ce papier est organisé comme suit : avant de présenter le modèle économique développant le choix de traitement suivant les différents modes de régulation (section 3), nous développons un exemple de biais d'endogénéité (section 2). Nous poursuivons ensuite par une présentation des données (section 4) qui nous permettent d'estimer et de montrer qu'il existe un biais d'endogénéité dans l'estimation des coûts de traitement

2 Approche simplifiée du problème

A travers l'exemple suivant, nous tentons de résumer l'idée principale développée dans ce papier. Supposons que deux traitements soient disponibles pour une même pathologie : L et H . Le traitement L est moins coûteux et convient mieux aux cas les moins graves. Le traitement H , quant à lui, est plus coûteux et est mieux adapté aux cas les plus sévères. Dans notre application empirique, le traitement L correspondra à un accouchement par voie basse et le traitement H à une césarienne. Supposons, que les patients soient répartis uniformément entre trois niveaux de gravité $(-1,0,+1)$, et que les coûts de traitement espérés pour chaque type de gravité suivant le choix de traitement sont donnés par le tableau suivant :

Gravité	Traitement L	Traitement H
-1	$\gamma_L - 1$	$\gamma_H - 1$
0	γ_L	γ_H
+1	$\gamma_L + 1$	$\gamma_H + 1$

On observe que pour tous les types de patients, le coût d'opportunité du traitement H est de $\gamma_H - \gamma_L$. Néanmoins, en pratique, on observe les coûts de traitement seulement *ex post*, c'est-à-dire après traitement. En excluant les cas où aucun des traitements n'est utilisé, et les cas où l'intensité des traitements ne serait pas croissante avec la gravité de la maladie, on constate qu'on peut se retrouver dans deux situations :

- Le traitement L est utilisé pour soigner seulement les patients de gravité -1, et le traitement H est utilisé pour les patients de types 0 et +1. Dans ce cas, le coût de traitement moyen observé *ex post* pour le traitement L est $\bar{c}_L = \gamma_L - 1$ et le coût moyen observé *ex post* pour le traitement H est $\bar{c}_H = \gamma_H + \frac{1}{2}$.
- Le traitement L est utilisé pour soigner les patients de gravité -1 et 0, et le traitement H est utilisé pour les patients de types +1. Dans ce cas, on a $\bar{c}_L = \gamma_L - \frac{1}{2}$ et de $\bar{c}_H = \gamma_H + 1$.

Dans ces deux cas, on constate que le coût d'opportunité du traitement H mesuré *ex post* sera de $\gamma_H - \gamma_L + \frac{3}{2}$, ce qui surestime évidemment le vrai coût d'opportunité.

On constate donc que si le régulateur propose un remboursement au niveau des coûts moyens pour chaque pathologie, l'hôpital fera un profit de $\frac{3}{2}$ par patient soigné avec le traitement H . Cette situation est moins efficace que le remboursement des coûts qui conduit à un profit nul de l'hôpital quel que soit le traitement effectué. Le remboursement au coût moyen est donc susceptible de déformer le choix de traitement. La section suivante, présente le modèle économique qui servira de base aux estimations économétriques.

3 Le modèle économique

3.1 Définitions et hypothèses

Nous nous intéressons à un modèle de choix de traitement dans lequel deux traitements (L et H) sont disponibles pour soigner la même pathologie. Le traitement L est un traitement léger, qui convient mieux au cas les moins graves. Le traitement H est plus coûteux et convient mieux aux cas les plus

graves. Le séjour de chaque patient est décrit par la réalisation d'un vecteur aléatoire $(\theta, x_1, \dots, x_K, s_L, s_H, C_L, C_H)$. Les composantes de ce vecteur aléatoire admettent les interprétations suivantes :

- θ représente la gravité de la maladie. On suppose que θ est distribué sur $]-\infty; +\infty[$ suivant une loi de densité f et de fonction de répartition F .
- $\mathbf{x} = (x_1, \dots, x_K)'$ représente un vecteur de K diagnostics.
- s_L (resp. s_H) représente le surplus que le traitement L (resp. H) apporte au patient.
- C_L (resp. C_H) représente un paramètre spécifique au patient influençant le coût du traitement L (resp. H). On nommera par la suite C_L ou C_H , le casemix du patient.

Les coûts de traitement sont également influencés par deux facteurs caractéristiques de l'hôpital dans lequel est soigné le patient :

- L'hôpital peut être plus coûteux qu'un autre établissement de façon exogène (i.e. hôpital urbain *vs.* hôpital rural ou hôpital universitaire *vs.* hôpital non universitaire). On mesurera ces disparités de coûts exogènes par un paramètre k caractéristique de l'hôpital. Ces disparités peuvent provenir de facteurs mesurables, comme de facteurs non mesurables. Cet aspect n'est pas fondamental dans notre étude, mais a été développé dans Choné et Lesur (2002).
- Les coûts de traitement peuvent être réduits grâce à un effort de réduction de coût $e \geq 1$ de la part du manager de l'hôpital. Cet effort est déterminé avant que les patients soient soignés et est identique pour tous les patients. Il engendre une désutilité de $\psi(e)$ pour le manager de l'hôpital. On supposera $\psi(1) = 0$, $\psi', \psi'' > 0$, et on définira $r = \psi'(1) > 0$.

Les coûts de traitement sont finalement donnés par :

$$c_L = \frac{k}{e}C_L, c_H = \frac{k}{e}C_H$$

Nous voyons que cette formalisation de l'effort de réduction de coût laisse inchangée la structure des coûts relatifs. La justification principale de cette fonction de coûts est qu'elle est celle sous-jacente à la régulation utilisée en pratique (voir McClellan [1997]).

La structure d'information est supposée être la suivante :

- Le patient ne connaît pas la réalisation du vecteur aléatoire, il sait seulement qu'il est malade.
- Le paramètre k ainsi que l'effort de l'hôpital e sont connus seulement par le manager de l'hôpital.

- Durant la période d’exercice, le médecin de l’hôpital observe seulement la réalisation de θ et de \mathbf{x} pour chaque patient. A partir de ces éléments, il détermine le traitement à appliquer à chaque patient.
- Le régulateur, qui est aussi l’économètre, observe *ex-post* le traitement réalisé (L ou H), la réalisation du vecteur \mathbf{x} et la réalisation du coût final c_L ou c_H suivant le choix de traitement qui aura été fait.
- La distribution du vecteur aléatoire $(\theta, x_1, \dots, x_K, s_L, s_H, C_L, C_H)$ et du modèle est connaissance commune.

Afin de mieux spécifier le modèle, nous aurons recours aux hypothèses suivantes :

H_Δ : Il existe une constante positive Δ telle que :

$$C_H = C_L + \Delta.$$

H_C : Quel que soit le traitement $t \in \{L, H\}$, le casemix espéré $E(C_t|\theta)$ est croissant avec la gravité.

H_s : Le surplus s apporté par le traitement H est strictement croissant avec la gravité θ : $s'(\theta) > 0$. De plus, il existe une valeur θ^* telle que $s(\theta^*) = 0$.

H_θ : Le paramètre de gravité θ est relié au vecteur \mathbf{x} : la loi de θ conditionnellement à \mathbf{x} est une loi normale de moyenne $\beta'\mathbf{x}$ où $\beta = (\beta_1, \dots, \beta_K)'$:

$$\theta = \beta'\mathbf{x} + u$$

avec $u \sim N(0, \sigma_u^2)$.

L’hypothèse H_Δ signifie que le traitement H est plus coûteux que le traitement L . L’hypothèse H_C est l’hypothèse centrale que nous souhaitons tester dans ce papier : il s’agit de savoir si le casemix du patient est lié à la gravité de celui-ci *indépendamment* du choix de traitement. L’hypothèse H_s paraît plutôt réaliste : elle signifie que le traitement lourd est mieux adapté aux cas les plus graves. Quant à l’hypothèse H_θ , elle sera surtout utile aux estimations économétriques du modèle, mais elle spécifie l’avantage informationnel du médecin face au régulateur : le médecin observe θ et \mathbf{x} , mais θ est plus informatif que \mathbf{x} . Le médecin se fondera donc uniquement sur l’observation de θ pour effectuer son choix de traitement. La variable aléatoire u capture donc toute l’information observée par le médecin et qui ne peut être retranscrit à travers le système d’information hospitalier.

Le traitement sera représenté par la variable $t \in \{L, H\}$. Nous utiliserons également la variable indicatrice d définie de la façon suivante :

$$\begin{aligned} d &= 1 \text{ si } t = H \\ d &= 0 \text{ si } t = L \end{aligned}$$

Comme nous l'avons vu, d'après H_θ , le médecin effectuera son choix de traitement en fonction de la réalisation de θ . On notera (abusivement) par $t(\theta)$ la fonction de choix de traitement du médecin, que nous déterminerons un peu plus loin. Nous supposons seulement que si le médecin est indifférent entre deux traitements, il choisit le traitement le moins coûteux, c'est-à-dire L .

Nous utiliserons également la fonction $s(\theta)$ qui représente l'espérance du surplus apporté par le traitement H par rapport au traitement L pour un patient de gravité θ :

$$s(\theta) = E(s_H - s_L | \theta).$$

En notant $p_H = 1 - p_L = Prob\{t = H\} = E(d)$, on peut écrire le surplus espéré du patient représentatif noté S :

$$S = Es_L + p_H E(s(\theta) | t(\theta) = H)$$

Le casemix final C est défini par :

$$C = (1 - d)C_L + dC_H$$

On en déduit immédiatement le casemix espéré \bar{C} :

$$\bar{C} = EC = p_L \bar{C}_L + p_H \bar{C}_H$$

avec $\bar{C}_L = E(C_L | t(\theta) = L)$ et $\bar{C}_H = E(C_H | t(\theta) = H)$. \bar{C}_L et \bar{C}_H représentent donc respectivement les casemix moyens des patients soignés avec les traitements L et H . Nous noterons les coûts moyens par des minuscules : $\bar{c} = \frac{k}{e} \bar{C}$, $\bar{c}_L = \frac{k}{e} \bar{C}_L$ et $\bar{c}_H = \frac{k}{e} \bar{C}_H$. \bar{c} représente le coût le traitement moyen dans l'établissement, \bar{c}_t représente donc respectivement le coût moyen du traitement t .

En utilisant l'hypothèse H_Δ , on obtient :

$$C = C_L + \Delta d.$$

On peut alors trouver une expression simple de l'espérance du casemix :

$$\bar{C} = E(C_L) + p_H \Delta$$

En notant $\Gamma_L = E(C_L)$, on peut écrire :

$$\bar{C} = \Gamma_L + p_H \Delta$$

Il est important de noter que Γ_L représente le casemix moyen du traitement L si tous les patients étaient soignés avec le traitement L . On notera par $\gamma_L = \frac{k}{e} \Gamma_L$ le coût moyen du traitement L si tous les patients étaient soignés avec le traitement L . On utilisera des notations équivalentes pour $\Gamma_H = \Gamma_L + \Delta$ et $\gamma_H = \frac{k}{e} \Gamma_H$. Ces notations sont bien évidemment à rapprocher avec celles utilisées dans l'exemple de la section 2.

3.2 L'objectif de l'hôpital

Nous supposons que l'hôpital prend ses décisions afin de maximiser un objectif U qui dépend du surplus des patients, de son profit et du coût lié à l'effort de réduction de coût. On notera le nombre de patients soignés dans l'hôpital par N . Plus précisément, on considérera que l'objectif de l'hôpital est de la forme :

$$U = q_1 NS + q_2 \Pi - \psi(e), \quad q_1, q_2 > 0$$

où Π représente le profit espéré de l'hôpital, qui dépend étroitement du mode de régulation utilisé. L'intérêt de cette formalisation est de traiter en toute généralité l'objectif d'un hôpital.

L'hôpital possède donc deux instruments sur lesquels il peut agir afin de maximiser son objectif : sa fonction de choix de traitement $t(\theta)$ et son effort de réduction de coût e .

Nous allons d'abord montrer qu'à partir de H_s et $q_1 > 0$, la fonction de choix de traitement $t(\theta)$ choisie par l'hôpital est monotone, à savoir qu'elle vérifie la propriété suivante : si $t(\theta) = H$ alors $t(\theta') = H$ pour tout $\theta' \geq \theta$. Remarquons tout d'abord que pour toute règle de traitement monotone, on peut définir de façon bijective un niveau $\tilde{\theta}$ tel que pour tout $\theta \leq \tilde{\theta}$ alors $t(\theta) = L$ et sinon $t(\theta) = H$.

La propriété de monotonie de la fonction de choix de traitement se montre alors comme suit : considérons une fonction de choix de traitement t quelconque et notons $p_H = \text{Prob}\{t(\theta) = H\}$. En utilisant H_s et $q_1 > 0$, la règle de traitement monotone définie par $\tilde{\theta} = F^{-1}(p_H)$ est nécessairement préférée par l'hôpital. Ainsi, toute fonction de choix de traitement est dominée par une fonction de choix de traitement monotone, et on peut donc restreindre notre analyse à ces seules fonctions.

Comme le médecin effectue un choix de traitement monotone, on obtient le résultat suivant :

$$\begin{aligned} \bar{C}_L &\leq \Gamma_L \\ \Gamma_L + \Delta &= \Gamma_H \leq \bar{C}_H \end{aligned}$$

Ces résultats proviennent des hypothèses H_C et H_Δ . De façon immédiate, on obtient :

$$\bar{c}_L \leq \gamma_L < \gamma_H \leq \bar{c}_H$$

Ce dernier résultat est à rapprocher de l'exemple de la section 2. L'objectif de ce papier est donc de tester si l'hypothèse H_C est vérifiée. Si le coût de traitement est strictement croissant avec la gravité, alors on a :

$$\bar{c}_L < \gamma_L < \gamma_H < \bar{c}_H$$

Par contre, si le coût de traitement est indépendant de la gravité, alors on aura :

$$\bar{c}_L = \gamma_L \text{ et } \bar{c}_H = \gamma_H$$

On constate donc le rôle crucial joué par l'hypothèse H_C et la nécessité de vérifier si elle est vérifiée empiriquement.

3.3 Le comportement de l'hôpital sous différents modes de régulation

L'objectif de cette sous-section est de présenter les différents modes de régulation hospitalière et d'en déduire le comportement de l'hôpital.

3.3.1 Le remboursement des coûts

Dans le cas où l'hôpital voit ses coûts remboursés, les transferts nets de coûts par patient sont nuls. Le profit espéré de l'hôpital est donc nul. L'objectif de l'hôpital se résume donc à la maximisation de $U = q_1 NS - \psi(e)$. Le premier terme de U ne dépend que du surplus des patients, c'est-à-dire du choix de traitement et le second terme ne dépend que de l'effort. On en déduit donc que l'effort e va être minimal, c'est-à-dire $e = 1$ et le choix de traitement va vérifier :

$$\begin{aligned} t(\theta) &= H \text{ ssi } s(\theta) > 0 \\ t(\theta) &= L \text{ ssi } s(\theta) \leq 0 \end{aligned}$$

Le choix de traitement est donc monotone et le niveau $\tilde{\theta}$ est donné par $s(\tilde{\theta}) = 0$, c'est-à-dire $\tilde{\theta} = \theta^*$. On constate que l'intérêt du remboursement des coûts est d'amener le surplus des patients au niveau maximal, ce qui n'est évidemment pas le niveau socialement optimal puisqu'aucun arbitrage avec le coût de traitement n'est pris en considération. L'inconvénient majeur de ce mode de remboursement est de ne pas inciter à l'effort de réduction de coûts.

3.3.2 Le budget global

L'hôpital reçoit un montant G pour financer son budget. Ce montant est indépendant du nombre de patients reçus et de la nature des traitements. Le profit de l'hôpital s'écrit alors :

$$\Pi \equiv G - N \frac{k}{e} (\Gamma_L + (1 - F(\tilde{\theta}))\Delta)$$

On peut donc écrire l'objectif de l'hôpital :

$$U = q_2 G + N q_1 E(s_L) + N q_1 \int_{\tilde{\theta}}^{+\infty} s(\theta) f(\theta) d\theta - N q_2 \frac{k}{e} (\Gamma_L + (1 - F(\tilde{\theta})) \Delta) - \psi(e)$$

Les dérivées partielles de U par rapport à $\tilde{\theta}$ et e s'expriment alors par :

$$\begin{aligned} \frac{\partial U}{\partial \tilde{\theta}} &= N \left(k q_2 \frac{\Delta}{e} - q_1 s(\tilde{\theta}) \right) f(\tilde{\theta}) \\ \frac{\partial U}{\partial e} &= N k q_2 \frac{\Gamma_L + (1 - F(\tilde{\theta})) \Delta}{e^2} - \psi'(e) \end{aligned}$$

Le problème du budget global est que les hôpitaux peuvent être rationnés. Pour prendre en compte ce phénomène, il faut rigoureusement écrire le programme de l'hôpital :

$$\begin{aligned} &\max_{\tilde{\theta}, e} U \\ &\text{s.c. } \Pi \geq 0 \quad (\lambda_1) \\ &\quad e \geq 1 \quad (\lambda_2) \end{aligned}$$

Les conditions du premier ordre conduisent à :

$$\begin{aligned} s(\tilde{\theta}) &= \frac{k}{e} \frac{(q_2 + \lambda_1)}{q_1} \Delta \\ \psi'(e) &= N (q_2 + \lambda_1) \frac{k}{e} \frac{\Gamma_L + (1 - F(\tilde{\theta})) \Delta}{e} + \lambda_2 \end{aligned}$$

Dans le cas du budget global, on constate tout d'abord que la valeur $\tilde{\theta}$ est supérieure à la valeur trouvée pour le remboursement des coûts. Cela signifie que sous le budget global, l'hôpital traite moins de patients avec le traitement H . Cela provient du fait qu'à travers le mécanisme du budget global, l'hôpital internalise en partie les coûts liés aux décisions thérapeutiques.

L'effort de réduction de coût peut être supérieur à 1, si cela est optimal pour l'hôpital. On supposera que cela est toujours vrai, pour ce faire, on admettra que l'hypothèse $r < N q_2 k \Gamma_L$ est toujours vérifiée.

On peut également s'intéresser à l'impact du rationnement ($\lambda_1 > 0$). La part de patients soignés avec le traitement H est décroissante avec λ_1 , ce qui est intuitif : si l'hôpital est rationné, alors il va soigner moins de patients avec le traitement coûteux afin de respecter sa contrainte budgétaire. On voit également que l'impact du rationnement sur l'effort n'est pas neutre : le manager de l'hôpital est incité à faire plus d'effort afin de relâcher le poids de la contrainte budgétaire.

3.3.3 La tarification à l'activité

La tutelle définit deux niveaux de casemix R_L et R_H qui seront remboursés à l'hôpital pour les traitements L et H . Afin de simplifier les mécanismes de paiement, on considérera que la tutelle prend en charge toute l'hétérogénéité des coûts des hôpitaux et qu'elle versera donc un montant $\frac{k}{e}R_L$ pour les traitements légers et un montant $\frac{k}{e}R_H$ pour rembourser les traitements lourds. Le profit espéré est donc :

$$\Pi = \frac{k}{e} \left(F(\tilde{\theta})(R_L - \bar{C}_L) + (1 - F(\tilde{\theta}))(R_H - \bar{C}_H) \right)$$

Nous supposons que le régulateur va proposer un contrat (R_L, R_H) telle que la contrainte de participation de l'hôpital va être vérifiée, c'est-à-dire $\Pi \geq 0$.

Après dérivation du programme de maximisation de l'hôpital, on obtient les variables d'effort et de choix de traitement suivantes :

$$\begin{aligned} \psi'(e) &= Nq_2 \frac{k}{e} \frac{\Gamma_L + (1 - F(\tilde{\theta}))\Delta}{e} \\ s(\tilde{\theta}) &= -\frac{q_2}{q_1} \frac{k}{e} [R_H - R_L - \Delta] \end{aligned}$$

On voit grâce à la seconde condition que $\tilde{\theta}$ dépend de la différence entre $R_H - R_L$ et Δ . Si $R_H - R_L$ est supérieur à Δ , alors on constate que le niveau $\tilde{\theta}$ est inférieur à θ^* , cela signifie que dans un tel cas, l'hôpital fera plus de traitement H que dans le cas du remboursement des coûts. De plus, un tel mode de remboursement ne maximise pas le surplus des patients : trop de traitements lourds sont effectués.

On peut remarquer qu'on retrouve les résultats obtenus avec le budget global dans le cas $R_L = R_H$. Cela est logique, car dans un tel cadre, le remboursement est alors identique pour tous les patients et est égal au budget moyen alloué à l'établissement.

Un des objectifs de la théorie de la régulation est de minimiser les rentes laissées aux firmes. Ici, pour que le profit soit nul, on voit qu'il faut que R_L et R_H vérifient :

$$\begin{aligned} R_L + (1 - F(\tilde{\theta}))(R_H - R_L) &= \Gamma_L + (1 - F(\tilde{\theta}))\Delta \\ \Leftrightarrow F(\tilde{\theta})R_L + (1 - F(\tilde{\theta}))R_H &= F(\tilde{\theta})\bar{C}_L + (1 - F(\tilde{\theta}))\bar{C}_H \end{aligned}$$

Il existe donc une infinité de contrats (R_L, R_H) permettant de vérifier $\Pi = 0$. Néanmoins, nous allons nous concentrer plus précisément sur deux d'entre eux :

Le contrat A Le contrat A va être fondé sur la pratique du régulateur observée en réalité : celui-ci fixe R_L et R_H au niveau des casemix moyens observés, c'est-à-dire $R_L = \overline{C}_L$, et $R_H = \overline{C}_H$. Chaque traitement est alors rentable. La logique de ce contrat est de rembourser chaque traitement au niveau du coût moyen des patients qui seront à l'aide de ce traitement.

Le contrat B Le contrat B sera celui défini par $R_L = \Gamma_L > \overline{C}_L$ et $R_H = \Gamma_L + \Delta < \overline{C}_H$. Ce contrat opère des subventions croisées entre traitements. La logique de ce contrat est de rembourser chaque traitement au niveau du coût moyen *si tous les patients étaient soignés* à l'aide de ce traitement.

Le contrat A est tel que $R_H - R_L > \frac{k}{e}\Delta$ et le contrat B est tel que $R_H - R_L = \Delta$. Ainsi, le contrat A mène à une utilisation du traitement H plus importante que pour le remboursement des coûts et le budget global. Ainsi, l'effet final sur les coûts est indéterminé, car le contrat A incite à un niveau d'effort plus important que le remboursement des coûts, mais incite également les hôpitaux à effectuer plus de traitements H .

Notre propos est que lorsque les autorités de régulation sont passés d'un remboursement des coûts à la tarification à l'activité définie à l'aide du contrat A, alors le niveau $\hat{\theta}$ a pu diminuer à cause de l'écart entre $R_H - R_L$ et Δ . Ceci nous semble être une nouvelle explication plausible de l'accroissement du casemix moyen observé en pratique. Qui plus est, le danger avec ce contrat est qu'il ne maximise pas le surplus des patients : certains patients ne nécessitant pas le traitement H se voient soignés avec.

Le contrat B que nous proposons permet de rétablir le même niveau de choix de traitement que pour le remboursement des coûts. Néanmoins, il implique des subventions croisées entre les deux traitements. Le biais d'endogénéité que nous nous proposons de tester économétriquement a donc des conséquences extrêmement importantes en terme de régulation : il implique de la part du régulateur un arbitrage entre éviter les subventions croisées et implémenter une régulation optimale. Le danger de l'utilisation des subventions croisées est qu'elle peuvent inciter les hôpitaux à écrémer et sélectionner les patients. Néanmoins, nous avons montré que s'interdire l'utilisation de subventions croisées entre traitements ne permet pas d'induire une régulation optimale. Comme nous l'avons déjà signalé, ce résultat est intimement lié à l'hypothèse H_C qui doit donc être testée empiriquement.

4 Les données

Les données sur lesquelles les estimations empiriques ont été réalisées concernent les accouchements ayant eu lieu en 1999 et 2000 dans des hôpitaux publics français appartenant à l'échantillon national de coûts (ENC). Nous disposons, pour chaque accouchement, du coût total du séjour mesuré grâce à la comptabilité analytique des établissements, ainsi que des variables caractéristiques du séjour : âge de la mère, ensemble des diagnostics (diagnostic principal et diagnostics associés), durée de séjour, présence de complications ou morbidités associées (CMA).

Les séjours conservés pour l'analyse concernent les femmes de 15 à 45 ans ayant effectué un séjour complet dans l'établissement (pas de transfert en entrée ou en sortie). Les séjours analysés appartiennent aux GHM 530 (césariennes avec CMA), 531 (césariennes sans CMA), 539 (accouchements par voie basse avec CMA) et 540 (accouchements par voie basse sans CMA) et proviennent de 26 hôpitaux publics pour 1999, et de 28 hôpitaux pour 2000. Les statistiques descriptives sur les coûts des accouchements dans notre échantillon sont présentées dans les tableaux 1 et 2.

Groupes Homogènes de Malades	<i>Analyse de la variance</i>			Moyenne relative*	Points ISA**
	Fréquence	Moyenne	Écart-type		
530 : César. avec CMA	2,85%	24.424	14.533	1.853	2.155
531 : Césariennes	12,93%	19.501	9.602	1.479	1.507
539 : V. basse avec CMA	12,30%	15.884	9.704	1.205	1.158
540 : Acc. par voie basse	71,92%	13.183	5.364	1.000	1.000
Ensemble	$N=39.111$	14.652	7.596		
		$R^2=0,13$			

*base 1.000=GHM540

**Source : Agence Technique de l'Information sur l'Hospitalisation

FIG. 1 – Coût (en francs courants) par GHM pour l'année 1999

La classification en GHM explique entre 11% et 13% de la variance des coûts de traitement observés. Il reste donc une forte variance au sein de chaque GHM, qui a pour origine des différences de productivité et d'efficacité entre établissements (le facteur $\frac{k}{e}$ dans le modèle théorique) ainsi que des différences de casemix au niveau des patientes.

On constate sur données brutes qu'une césarienne sans CMA coûtait, en 1999, 47,9% plus cher qu'un accouchement par voie basse sans CMA. Ce différentiel de coût est de 44% en 2000. Néanmoins, calculer des moyennes de coût sur données brutes ne donne qu'une grossière idée de la différence réelle, car la productivité des hôpitaux peut être variable et le coût relatif

Groupes Homogènes de Malades	<i>Analyse de la variance</i>			Moyenne relative*	Points ISA **
	Fréquence	Moyenne	Écart-type		
530 : César. avec CMA	2,77%	26.783	20.285	1.952	2.229
531 : Césariennes	14,44%	19.755	10.980	1.440	1.431
539 : V. basse avec CMA	12,95%	15.842	10.527	1.155	1.160
540 : Acc. par voie basse	69,84%	13.718	5.655	1.000	1.000
Ensemble	$N=43.974$	15.226	8.589		
		$R^2=0,11$			

*base 1.000=GHM540

**Source : Agence Technique de l'Information sur l'Hospitalisation

FIG. 2 – Coût (en francs courants) par GHM pour l'année 2000

de la césarienne peut donc varier d'un hôpital à l'autre. Il conviendra donc de déterminer précisément le coût relatif d'une césarienne en neutralisant ces différentiels de productivité entre établissements.

La proportion de césariennes dans notre échantillon est de 15,78% en 1999 et de 17,21% en 2000. Ces nombres sont à comparer avec l'ensemble des accouchements ayant eu lieu dans le secteur hospitalier en France ces mêmes années (voir tableaux 3 et 4). Il semble d'après le tableau 3 que notre

Groupes Homogènes de Malades	ENC (rappel)	Base Publique	Base Privée	Ensemble
530 : César. avec CMA	2,85%	2,45%	1,63%	2,17%
531 : Césariennes	12,93%	13,36%	15,55%	14,11%
539 : V. basse avec CMA	12,30%	12,29%	8,22%	10,89%
540 : Acc. par voie basse	71,92%	71,90%	74,60%	72,83%
Total	39.111	464.997	243.857	708.854
<i>Taux de césariennes</i>	15,78%	15,81%	17,18%	16,28%

FIG. 3 – Casemix français en 1999

échantillon soit, en terme de taux de césariennes, assez bien représentatif des hôpitaux publics pour l'année 1999. Néanmoins, la représentativité de l'échantillon sur l'année 2000 est moins assurée. Le biais éventuel que pourrait présenter notre échantillon n'est pas gênant quant au test de l'hypothèse H_C que nous nous sommes proposés de faire. Néanmoins, les simulations de politique économique que nous effectuons par la suite peuvent être influencées par la représentativité de nos données.

Groupes Homogènes de Malades	ENC (rappel)	Base Publique	Base Privée	Ensemble
530 : César. avec CMA	2,77%	2,76%	1,67%	2,38%
531 : Césariennes	14,44%	13,90%	16,07%	14,66%
539 : V. basse avec CMA	12,95%	13,37%	8,30%	11,58%
540 : Acc. par voie basse	69,84%	69,97%	73,96%	71,38%
Total	43.974	504.572	274.056	778.628
<i>Taux de césariennes</i>	<i>17,21%</i>	<i>16,65%</i>	<i>17,75%</i>	<i>17,04%</i>

FIG. 4 – Casemix français en 2000

5 Spécification du modèle

Afin de pouvoir expliciter le comportement de l'hôpital, nous supposons par la suite que le coût fixe lié à l'effort de réduction de coût a la forme suivante :

$$\psi(e) = \frac{r}{\alpha}(e^\alpha - 1)$$

avec $\alpha > 1$. Cette spécification vérifie bien les hypothèses sur ψ , à savoir : $\psi(1) = 0$, $\psi'(1) = r$, et $\psi', \psi'' > 0$.

Plusieurs formalisations respectant les hypothèses H_Δ et H_C sont possibles. Nous développerons par la suite une seule d'entre elle afin de l'estimer économétriquement. On suppose que le casemix C a la forme suivante :

$$C = a + \Delta d + \rho\theta + v$$

où la variable aléatoire v suit une loi normale $N(0, \sigma_v^2)$. La covariance entre u et v est notée σ_{uv} et v est supposé être non corrélé avec \mathbf{x} .

L'hypothèse H_θ permet d'écrire :

$$\Gamma_L = a + \rho E(\theta) = a + \rho E_{\mathbf{x}}(\beta' \mathbf{x})$$

Le problème principal qu'on peut rencontrer avec une telle formalisation est que le coût de traitement peut théoriquement prendre des valeurs négatives. Néanmoins, nous montrerons qu'avec nos estimations, la probabilité pour que le coût C prenne des valeurs négatives est faible et que l'utilisation d'une loi normale ne pose pas de problème particulier.

En notant ϕ et Φ la densité et la fonction de répartition de la loi normale

centrée réduite, on peut exprimer les valeurs de \bar{C}_L et \bar{C}_H :

$$\bar{C}_L = a + \rho E_{\mathbf{x}}(\beta' \mathbf{x} | t = L) - \left(\rho \sigma_u + \frac{\sigma_{uv}}{\sigma_u}\right) E_{\mathbf{x}} \left(\frac{\phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)}{1 - \Phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)} \middle| t = L \right)$$

$$\bar{C}_H = a + \Delta + \rho E_{\mathbf{x}}(\beta' \mathbf{x} | t = H) + \left(\rho \sigma_u + \frac{\sigma_{uv}}{\sigma_u}\right) E_{\mathbf{x}} \left(\frac{\phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)}{\Phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)} \middle| t = H \right)$$

On en déduit :

$$\begin{aligned} \bar{C}_H - \bar{C}_L = & \Delta + \rho [E_{\mathbf{x}}(\beta' \mathbf{x} | t = H) - E_{\mathbf{x}}(\beta' \mathbf{x} | t = L)] \\ & + \left(\rho \sigma_u + \frac{\sigma_{uv}}{\sigma_u}\right) \left[E_{\mathbf{x}} \left(\frac{\phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)}{\Phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)} \middle| t = H \right) + E_{\mathbf{x}} \left(\frac{\phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)}{1 - \Phi\left(\frac{\beta' \mathbf{x} - \tilde{\theta}}{\sigma_u}\right)} \middle| t = L \right) \right] \end{aligned}$$

Afin que l'hypothèse H_C soit vérifiée, il faut avoir $\rho > 0$ ou $\eta > 0$. Nous pourrions conclure à l'absence de biais d'endogénéité si $\rho = \eta = 0$. Néanmoins, si ρ ou η est strictement positif, nous aurons donc un problème d'endogénéité et la différence entre les coûts moyens surestimera le coût d'opportunité du traitement.

Concernant la fonction de surplus des patients, nous allons nous baser sur les recommandations conjointes de l'OMS, de l'UNESCO et de l'UNFPA : il est estimé que le taux de césariennes dans une population ne devrait pas être inférieur à 5%. Néanmoins, il est très difficile d'établir une limite supérieure pour ce taux de césarienne, étant donné qu'il est lié aux pratiques médicales, ainsi que des demandes des patientes qui peuvent évoluer dans le temps. Dans nos simulations de politiques économiques, nous présenterons donc nos résultats pour plusieurs valeurs de θ^* . Nous considérerons une fonction de surplus vérifiant les hypothèses suivantes :

$$\begin{aligned} \lim_{\theta \rightarrow -\infty} s(\theta) &= -\infty \\ s(\theta^*) &= 0 \\ \lim_{\theta \rightarrow \theta_{95}^-} s(\theta) &= +\infty \end{aligned}$$

où θ_{95} représente le 95^{ème} percentile de la distribution des θ , et θ^* représente le niveau de gravité d'indifférence entre césarienne et non-césarienne.

Plus précisément, nous utiliserons la fonction suivante vérifiant cette propriété :

$$s(\theta) = \kappa \cdot \ln \left(\frac{\theta_{95} - \theta^*}{\theta_{95} - \theta} \right), \quad \kappa > 0.$$

6 Présentation des méthodes d'estimation

Les données que nous avons portent sur les patientes hospitalisées pour un accouchement. Nos données portent donc sur H hôpitaux avec $h \in \{1, \dots, H\}$ représentant l'indice de l'hôpital. Nous noterons par N_h le nombre d'accouchements réalisés par l'hôpital h . Au sein d'un hôpital h , nous indexerons les patientes par $i \in \{1, \dots, N_h\}$. Le nombre total d'observations sera noté N avec $N = \sum_h N_h$.

A l'aide de ces notations, nous pouvons écrire le modèle économétrique correspondant aux réalisations du modèle économique présenté dans la section précédente :

$$\begin{aligned} c_{hi} &= \frac{k_h}{e_h} (a + \Delta d_{hi} + \rho \theta_{hi} + v_{hi}) \\ \theta_{hi} &= \beta' \mathbf{x}_{hi} + u_{hi} \\ \text{avec } \begin{cases} d_{hi} = 1 & \text{si } \theta_{hi} > \tilde{\theta}_h \\ d_{hi} = 0 & \text{si } \theta_{hi} \leq \tilde{\theta}_h \end{cases} \end{aligned}$$

Les vecteurs (u_{hi}, v_{hi}) suivent une loi normale bivariée vérifiant les hypothèses suivantes :

$$\begin{aligned} E(u_{hi}) &= E(v_{hi}) = 0 \\ E(u_{hi}^2) &= \sigma_u^2 \\ E(v_{hi}^2) &= \sigma_v^2 \\ E(u_{hi}v_{hi}) &= \sigma_{uv} \\ E(u_{hi}v_{h'i'}) &= E(u_{hi}u_{h'i'}) = E(v_{hi}v_{h'i'}) = 0 \text{ pour } (h, i) \neq (h', i') \end{aligned}$$

La forme semi-réduite est donc :

$$\begin{aligned} c_{hi} &= \frac{k_h}{e_h} (a + \Delta d_{hi} + \rho(\beta' \mathbf{x}_{hi}) + \varepsilon_{hi}) \\ d_{hi}^* &= \beta' \mathbf{x}_{hi} - \tilde{\theta}_h + u_{hi} \end{aligned}$$

avec $\varepsilon_{hi} = \rho u_{hi} + v_{hi}$. On a donc : $\sigma_{\varepsilon} = \rho \sigma_u^2 + \sigma_{uv}$ et $\sigma_{\varepsilon}^2 = \rho^2 \sigma_u^2 + \sigma_v^2 + 2\sigma_{uv}$. De plus, on a

$$\begin{aligned} d_{hi} &= 1 \text{ si } d_{hi}^* > 0 \\ d_{hi} &= 0 \text{ si } d_{hi}^* \leq 0 \end{aligned}$$

Le système précédent étant triangulaire (d_{hi} n'intervient pas dans l'équation définissant d_{hi}^*), on peut donc l'estimer. Toutefois, le modèle n'est pas identifiable en tant que tel. On posera par la suite les restrictions identifiantes : $a = \sigma_u = 1$.

Cette forme semi-réduite a comme propriété que le vecteur $(\varepsilon_{hi}, u_{hi})$ suit une loi normale bivariée. En notant par η le coefficient de corrélation reliant ε_{hi} et u_{hi} , on peut écrire : $\varepsilon_{hi} = \eta \frac{\sigma_\varepsilon}{\sigma_u} u_{hi} + e_{hi}$, avec par construction u_{hi} et e_{hi} indépendants. On propose trois stratégies d'estimation :

- une estimation en deux étapes de type Heckit (Probit suivi de NLS)
- une estimation en deux étapes par maximum de vraisemblance à information limitée (LIML)
- une estimation par maximum de vraisemblance à information complète (FIML)

6.1 Estimation en deux étapes : Probit suivi de NLS

En suivant la méthode de résolution en deux étapes à la Heckman, on peut écrire les espérances conditionnelles de c_{hi} :

$$E(c_{hi} | d_{hi} = 0, \mathbf{x}_{hi}) = \frac{k_h}{e_h} \left(1 + \rho(\beta' \mathbf{x}_{hi}) - \eta \sigma_\varepsilon \frac{\phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)}{1 - \Phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)} \right)$$

De même, on trouve :

$$E(c_{hi} | d_{hi} = 1, \mathbf{x}_{hi}) = \frac{k_h}{e_h} \left(1 + \Delta + \rho(\beta' \mathbf{x}_{hi}) + \eta \sigma_\varepsilon \frac{\phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)}{\Phi(\beta' \mathbf{x}_{hi} - \tilde{\theta}_h)} \right)$$

Les deux étapes de la résolution sont alors les suivantes :

- On fait tout d'abord un probit sur le choix de traitement. On obtient ainsi des estimations $\hat{\beta}$ et \hat{d}_{hi}^* de β et d_{hi}^* .
- En posant $Z_{hi}^1 = \hat{\beta}' \mathbf{x}_{hi}$ et $Z_{hi}^2 = d_{hi} \cdot \frac{\phi(\hat{d}_{hi}^*)}{\Phi(\hat{d}_{hi}^*)} - (1 - d_{hi}) \frac{\phi(\hat{d}_{hi}^*)}{1 - \Phi(\hat{d}_{hi}^*)}$, on estime alors par Moindres Carrés Non Linéaires (NLS) l'équation suivante :

$$c_{hi} = \frac{k_h}{e_h} (1 + \Delta d_{hi} + \rho Z_{hi}^1 + (\eta \sigma_\varepsilon) Z_{hi}^2) + \varepsilon'_{hi}$$

L'inconvénient avec cette méthode est que les résidus sont hétéroscédastiques. Afin de prendre en compte ce problème, on se tourne donc vers des méthodes fondées sur la vraisemblance du modèle.

6.2 Estimation par maximum de vraisemblance à information limitée (LIML)

La log-vraisemblance s'écrit :

$$\begin{aligned} \ln L = & -\frac{N}{2} \ln 2\pi - N \ln(\sigma_\varepsilon) - \sum_h N_h \ln\left(\frac{k_h}{e_h}\right) \\ & + \sum_{h,i} \left[\ln \Phi\left(\omega_{hi} \frac{\eta \frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \beta' \mathbf{x}_{hi} - \tilde{\theta}_h}{\sqrt{1-\eta^2}}\right) - \frac{1}{2} \left(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}\right)^2 \right] \end{aligned}$$

avec $\varepsilon_{hi} = \frac{e_h}{k_h} c_{hi} - (1 + \Delta d_{hi} + \rho \beta' \mathbf{x}_{hi})$ et $\omega_{hi} = 2d_{hi} - 1$.

Démonstration : On réécrit le modèle :

$$\begin{aligned} c_{hi} &= \frac{k_h}{e_h} (1 + \Delta d_{hi} + \rho(\beta' \mathbf{x}_{hi}) + \varepsilon_{hi}) \\ d_{hi}^* &= \beta' \mathbf{x}_{hi} - \tilde{\theta}_h + u_{hi} \end{aligned}$$

Les vecteurs $(\varepsilon_{hi}, u_{hi})$ suivent une loi normale bivariée vérifiant les hypothèses suivantes :

$$\begin{aligned} E(\varepsilon_{hi}) &= E(u_{hi}) = 0 \\ E(\varepsilon_{hi}^2) &= \sigma_\varepsilon^2 \\ E(u_{hi}^2) &= 1 \\ E(\varepsilon_{hi} u_{hi}) &= \eta \sigma_\varepsilon \\ E(\varepsilon_{hi} u_{h'i'}) &= E(\varepsilon_{hi} \varepsilon_{h'i'}) = E(u_{hi} u_{h'i'}) = 0 \text{ pour } (h, i) \neq (h', i') \end{aligned}$$

La densité de la loi normale bivariée (ε, u) est donnée par :

$$\frac{1}{\sigma_\varepsilon} \phi_2\left(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta\right) = \frac{1}{2\pi\sigma_\varepsilon\sqrt{1-\eta^2}} e^{-\frac{1}{2} \frac{\frac{\varepsilon^2}{\sigma_\varepsilon^2} + u^2 - 2\eta \frac{\varepsilon u}{\sigma_\varepsilon}}{1-\eta^2}}$$

A partir de l'expression précédente, on trouve les égalités suivantes :

$$\int_{-\infty}^{\bar{u}} \frac{1}{\sigma_\varepsilon} \phi_2\left(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta\right) du = \frac{1}{\sigma_\varepsilon} \phi\left(\frac{\varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\bar{u} - \eta \frac{\varepsilon}{\sigma_\varepsilon}}{\sqrt{1-\eta^2}}\right)$$

et

$$\int_{\underline{u}}^{+\infty} \frac{1}{\sigma_\varepsilon} \phi_2\left(\frac{\varepsilon}{\sigma_\varepsilon}, u, \eta\right) du = \frac{1}{\sigma_\varepsilon} \phi\left(\frac{\varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\eta \frac{\varepsilon}{\sigma_\varepsilon} - \underline{u}}{\sqrt{1-\eta^2}}\right)$$

La vraisemblance pour l'observation (h, i) s'écrit donc :

$$\begin{aligned} L &= \frac{1}{\frac{k_h}{e_h} \sigma_\varepsilon} \phi\left(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}\right) \left[\Phi\left(\frac{\tilde{\theta}_h - \beta' \mathbf{x}_{hi} - \eta \frac{\varepsilon_{hi}}{\sigma_\varepsilon}}{\sqrt{1 - \eta^2}}\right) \right]^{1-d_{hi}} \left[\Phi\left(\frac{\eta \frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \beta' \mathbf{x}_{hi} - \tilde{\theta}_h}{\sqrt{1 - \eta^2}}\right) \right]^{d_{hi}} \\ &= \frac{1}{\frac{k_h}{e_h} \sigma_\varepsilon} \phi\left(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}\right) \Phi\left(\omega_{hi} \frac{\eta \frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \beta' \mathbf{x}_{hi} - \tilde{\theta}_h}{\sqrt{1 - \eta^2}}\right) \end{aligned}$$

où $\omega_{hi} = 2d_{hi} - 1$. **CQFD**

La méthode LIML procède en deux étapes :

- La première étape est identique à la méthode précédente : on fait d'abord un probit sur le choix de traitement. On obtient ainsi des estimations $\hat{\beta}$ et \hat{d}_{hi}^* de β et d_{hi}^* .
- On maximise la vraisemblance suivante en $(\frac{k_h}{e_h}, \Delta, \rho, \eta, \sigma_\varepsilon)$:

$$\begin{aligned} \ln L &= -\frac{N}{2} \ln 2\pi - N \ln(\sigma_\varepsilon) - \sum_h N_h \ln\left(\frac{k_h}{e_h}\right) \\ &\quad + \sum_{h,i} \left[\ln \Phi\left(\omega_{hi} \frac{\eta \frac{\varepsilon_{hi}}{\sigma_\varepsilon} + \hat{d}_{hi}^*}{\sqrt{1 - \eta^2}}\right) - \frac{1}{2} \left(\frac{\varepsilon_{hi}}{\sigma_\varepsilon}\right)^2 \right] \end{aligned}$$

avec $\varepsilon_{hi} = \frac{e_h}{k_h} c_{hi} - (1 + \Delta d_{hi} + \rho \beta' \mathbf{x}_{hi})$ et $\omega_{hi} = 2d_{hi} - 1$.

L'avantage de la méthode LIML est que la vraisemblance est maximisée sur un nombre restreint de paramètres.

6.3 Estimation par maximum de vraisemblance à information complète (FIML)

Cette dernière méthode consiste à maximiser la vraisemblance du modèle sur tous les paramètres, à savoir le vecteur $(\frac{k_h}{e_h}, \Delta, \rho, \eta, \sigma_\varepsilon, \tilde{\theta}_h, \beta)$.

7 Résultats

Les trois estimations ont été réalisées sur chacune des deux années. La description du vecteur \mathbf{x} est consignée en annexe.

7.1 Confirmation du biais d'endogénéité

Les résultats du tableau 5 confirment l'existence d'un biais d'endogénéité. On constate que l'estimateur LIML n'est pas très performant. Les estimateurs Heckit et FIML mènent à des ordres de grandeur semblables et confirment l'existence d'un biais d'endogénéité.

	Année 1999			Année 2000		
	2-Step	LIML	FIML	2-Step	LIML	FIML
ρ	0.0516**	0.0219**	0.0557**	0.0338**	-0.0018	0.0245**
η	-	-0.0141	0.0414**	-	-0.0223*	0.0343**
σ_ε	-	0.4443**	0.4497**	-	0.4422**	0.4495**
$\eta\sigma_\varepsilon$	0.0580**	-	-	0.0352**	-	-

** sign. à 1% * sign. à 5%

FIG. 5 – Résultats concernant le biais d'endogénéité

7.2 L'hétérogénéité entre les choix de traitement

Présenter les résultats sur le lien entre $\tilde{\theta}_h$ et $\frac{k_h}{e_h}$.

En annexe : β

Conséquences en termes de régulation : forme des contrats A et B sans prendre en compte la modification de choix de traitement induite (ex si $q_2 = \lambda_1 = 0$ donc $e = 1$). C'est-à-dire toutes les différences inter-hôpitaux en terme de choix de traitement sont expliquées par des facteurs exogènes : s différent suivant les hôpitaux (maîtrise différente des technologies).

Contrat	Voie basse	Césarienne
A	1.000	1.482
B	1.000	1.377

Contrat	Voie basse sans CMA (540)	Voie basse avec CMA (539)	Césarienne sans CMA (531)	Césarienne avec CMA (530)
A	1.000	1.201	1.469	1.786
B	1.000	1.191	1.364	1.648

7.3 Simulation de politiques économiques

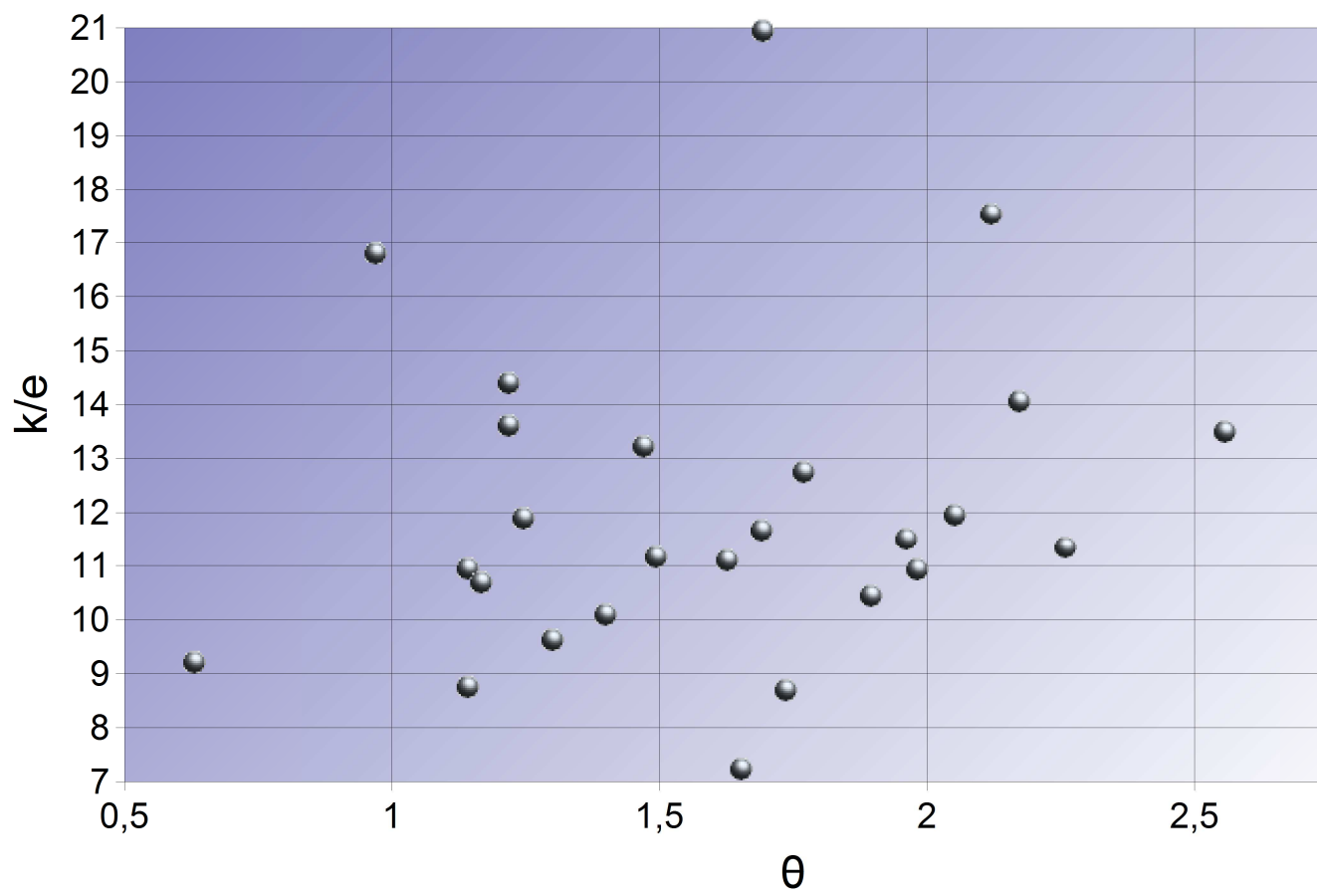
Les hôpitaux concernés par notre étude étaient tous financés sous le mode du budget global. On peut donc réécrire les résultats de la section 3.3.2, en supposant que tous les hôpitaux font un effort supérieur à 1 et qu'ils ont la même fonction objectif :

$$\ln \left(\frac{\theta_{95} - \theta^*}{\theta_{95} - \tilde{\theta}_h} \right) = \frac{k_h}{e_h} \Delta \frac{q_2 + \lambda_h}{q_1 \kappa}$$

$$\frac{r}{q_1 \kappa} e_h^\alpha = N_h \frac{k_h}{e_h} (\Gamma_L + (1 - F(\tilde{\theta}_h)) \Delta) \frac{q_2 + \lambda_h}{q_1 \kappa}$$

Estimation de $\tilde{\theta}_h$: Nous avons décidé d'utiliser une estimation de $\tilde{\theta}_h$ à l'aide de la distribution estimée de θ pour chaque hôpital. Une alternative aurait été

Répartition des hôpitaux en 1999



d'utiliser l'estimateur de $\tilde{\theta}_h$ du probit. Néanmoins, ceux-ci présente le même inconvénient que l'estimateur de la moyenne dans un probit (voir Greene). Ce problème est classique et décrit dans Greene. La solution proposée par ce dernier, et que nous utilisons est donc de prendre la proportion observée de césariennes, et d'estimer $\tilde{\theta}_h$ par fractile.

La démarche adoptée est donc la suivante : on se donne une valeur de α et de θ^* . Les étapes sont ensuite les suivantes :

1. On estime $\frac{q_2 + \lambda_h}{q_1 \kappa}$ avec la première équation.
2. On estime le min $\frac{q_2 + \lambda_h}{q_1 \kappa}$. En faisant l'hypothèse que dans notre échantillon, il existe un hôpital non rationné, on obtient de cette façon une estimation de $\frac{q_2}{q_1 \kappa}$.
3. On estime $\frac{r}{q_1 \kappa} e_h^\alpha$ avec la seconde équation.
4. Grâce à l'estimation de $\frac{r}{q_1 \kappa} e_h^\alpha$ et de $\frac{k_h}{e_h}$, on peut déduire une estimation de $(\frac{r}{q_1 \kappa})^{\frac{1}{\alpha}} k_h$.

Les étapes de la simulation de la tarification à la pathologie sont alors les suivantes :

1. Simulation du contrat B : on sait que c'est la valeur de θ^* qui sera choisie pour $\tilde{\theta}$. On sait donc que l'effort l'hôpital sera donné par

$$\left(\frac{r}{q_1 \kappa} e_h^\alpha \right)^{\frac{1+\alpha}{\alpha}} = N_h \left[\left(\frac{r}{q_1 \kappa} \right)^{\frac{1}{\alpha}} k_h \right] (\Gamma_L + (1 - F(\theta^*)) \Delta) \frac{q_2}{q_1 \kappa}$$

2. Simulation du contrat A : le problème principal est que $\tilde{\theta}_h$ et e_h sont solution implicite du système :

$$\ln \left(\frac{\theta_{95} - \theta^*}{\theta_{95} - \tilde{\theta}_h} \right) = - \left(R_H - R_L - \frac{k_h}{e_h} \Delta \right) \frac{q_2}{q_1 \kappa}$$

$$\frac{r}{q_1 \kappa} e_h^\alpha = N_h \frac{k_h}{e_h} (\Gamma_L + (1 - F(\tilde{\theta}_h)) \Delta) \frac{q_2}{q_1 \kappa}$$

avec R_H et R_L donnés par l'estimation du coût moyen, et dépendent donc directement de $\tilde{\theta}_h$. La stratégie d'estimation est alors la suivante : on effectue un balayage des niveaux $\tilde{\theta}_h$. Pour chaque valeur, on calcule deux estimations de $\frac{k_h}{e_h}$ grâce à chacune des deux équations. Le niveau $\tilde{\theta}_h$ qui minimise alors la distance entre les deux estimations de $\frac{k_h}{e_h}$ est alors retenue.

Suite des travaux économétriques Les résultats des simulations ne sont pas toutes présentées ici. Elles seront bien évidemment présentes dans la version finale. De même, une conclusion et une discussion des résultats sera présentée.

A Annexe : Liste des variables caractérisant le séjour

Code Libellé

Variables descriptives du séjour

AGE	Age de la patiente
CMA	Indicatrice de présence d'au moins une CMA
MULTI	Indicatrice de naissances multiples

Indicatrices de présence d'au moins un diagnostic dans les catégories suivantes de la CIM10

B95	Agents bactériens, cause de maladies classées dans d'autres chapitres
O13	Hypertension gestationnelle [liée à la grossesse] (sans protéinurie importante)
O14	Hypertension gestationnelle [liée à la grossesse] avec protéinurie importante
O22	Complications veineuses de la grossesse
O23	Infections de l'appareil génito-urinaire au cours de la grossesse
O24	Diabète sucré au cours de la grossesse
O26	Soins maternels pour d'autres affections liées principalement à la grossesse
O32	Soins maternels pour présentation anormale connue ou présumée du fœtus
O33	Soins maternels pour disproportion foeto-pelvienne connue ou présumée
O34	Soins maternels pour anomalie connue ou présumée des organes pelviens
O36	Soins maternels pour d'autres affections connues ou présumées du fœtus
O41	Autres anomalies du liquide amniotique et des membranes
O42	Rupture prématurée des membranes
O48	Grossesse prolongée
O60	Accouchement avant terme
O61	Échec du déclenchement du travail
O62	Anomalies de la contraction utérine et de la dilatation du col
O64	Dystocie due à une position et une présentation anormales du fœtus
O65	Dystocie due à une anomalie pelvienne de la mère
O66	Autres dystocies
O68	Travail et accouchement compliqués d'une détresse foetale
O71	Autres traumatismes obstétricaux
O75	Autres complications du travail et de l'accouchement, non classées ailleurs
O86	Autres infections puerpérales
O87	Complications veineuses au cours de la puerpéralité
O92	Autres mastopathies et anomalies de la lactation associées à l'accouchement
O98	Maladies infectieuses et parasitaires de la mère classées ailleurs mais compliquant la grossesse, l'accouchement et la puerpéralité
Z29	Nécessité d'autres mesures prophylactiques
Z35	Surveillance d'une grossesse à haut risque
Z86	Antécédents personnels de certaines autres maladies
Z87	Antécédents personnels d'autres maladies
Z88	Antécédents personnels d'allergie à des médicaments et des substances biologiques