

Direction de la recherche, des études,  
de l'évaluation et des statistiques  
DREES

SERIE  
SOURCES ET METHODES

**DOCUMENT  
DE  
TRAVAIL**

Calage sur marges de la base  
des cliniques privées pour améliorer  
l'estimation de la situation économique

Johanne AUDE

n° 17 – octobre 2010

MINISTERE DU TRAVAIL, DE LA SOLIDARITE ET DE LA FONCTION PUBLIQUE  
MINISTERE DE LA SANTE ET DES SPORTS  
MINISTERE DU BUDGET, DES COMPTES PUBLICS ET DE LA REFORME DE L'ÉTAT

## SOMMAIRE

<b>I. Qui sont les cliniques non répondantes ?</b> .....	<b>6</b>
<b>II. Le redressement de la non-réponse</b> .....	<b>8</b>
1. Aspects théoriques du calage sur marges .....	8
2. Le calage sur marges utilisé .....	10
a) Le choix des variables auxiliaires .....	10
b) La méthode de calage retenue .....	12
<b>III. Résultats du calage</b> .....	<b>13</b>
<b>IV. Quelques résultats de l'étude</b> .....	<b>14</b>
<b>Conclusion</b> .....	<b>17</b>

La DREES établit annuellement un bilan de la situation économique et financière des cliniques privées à but lucratif à partir de la base de données DIANE, constituée par la COFACE à partir des comptes que ces établissements ont l'obligation de déposer auprès des tribunaux de commerce. Cette base de données, qui regroupe les liasses fiscales de l'ensemble des entreprises françaises ayant déposé leur compte, fournit les données économiques et financières ; elle recueille et analyse les informations générales et les comptes sociaux des entreprises. La base de données comporte diverses informations sur les cliniques, telles que la raison sociale, le numéro SIREN, le statut juridique et le secteur d'activité de l'entreprise, l'actionnariat et les filiales ainsi que les différents postes du bilan et du compte de résultat. En revanche, la base ne permet de traiter que les comptes des sociétés d'exploitation des cliniques privées, car aucune méthodologie n'a permis jusqu'à présent de reconstituer les liens entre la société d'exploitation d'une clinique et les autres sociétés civiles, y compris les éventuelles SCI qui y sont rattachées<sup>1</sup>.

Par ailleurs, 20 % des cliniques privées à but lucratif ne déposent pas leurs comptes, et pour ces dernières, aucune information économique et financière n'est disponible. Utiliser les données disponibles dans DIANE sans correction préalable de la non-réponse totale est susceptible d'affecter la robustesse des indicateurs estimés, notamment lorsqu'il s'agit de calculer des évolutions. Il est en effet probable que les établissements qui ne déposent pas leurs comptes ont des caractéristiques particulières et ne sont pas représentatives de l'ensemble des cliniques privées.

Nous étudions dans un premier temps les caractéristiques des établissements qui ne déposent pas leurs comptes en rapprochant les données de DIANE de celles de la Statistique annuelle des établissements de santé (SAE), l'enquête annuelle exhaustive réalisée par la DREES auprès de l'ensemble des établissements de santé publics et privés<sup>2</sup>. Nous présentons ensuite la méthode de redressement de la non-réponse utilisée dans l'étude, le calage sur marges, qui permettra d'améliorer la qualité des résultats. Dans une troisième partie, nous analyserons comment le calage modifie les poids de sondage des cliniques afin de les rendre représentatifs de l'ensemble des établissements (et non plus des seuls répondants). Enfin, nous présenterons quelques résultats de l'étude, que nous comparerons avec ceux obtenus sans redresser la non-réponse.

\*  
\* \*

Les établissements étudiés sont ceux définis comme « cliniques privées à but lucratif » dans la SAE et classés dans les « activités hospitalières » par la nomenclature des activités françaises (NAF).

La plupart des cliniques présentes dans Diane 2008 pratiquent des « activités hospitalières » selon la NAF, néanmoins, 5 % d'entre elles appartiennent à des entités qui ne sont pas uniquement constituées de cliniques, tels que des groupes de gestion de maisons de retraite.

---

<sup>1</sup> Cf. Le Rhun B., Legendre M.-C., 2007, « L'évolution de la situation économique et financière des cliniques privées entre 2004 et 2005 », *Études et Résultats*, n° 583, DREES, juillet.

<sup>2</sup> La SAE permet notamment de connaître la discipline pratiquée par la clinique, son activité (entrées et venues) et ses facteurs de production associés (lits et places, personnel).

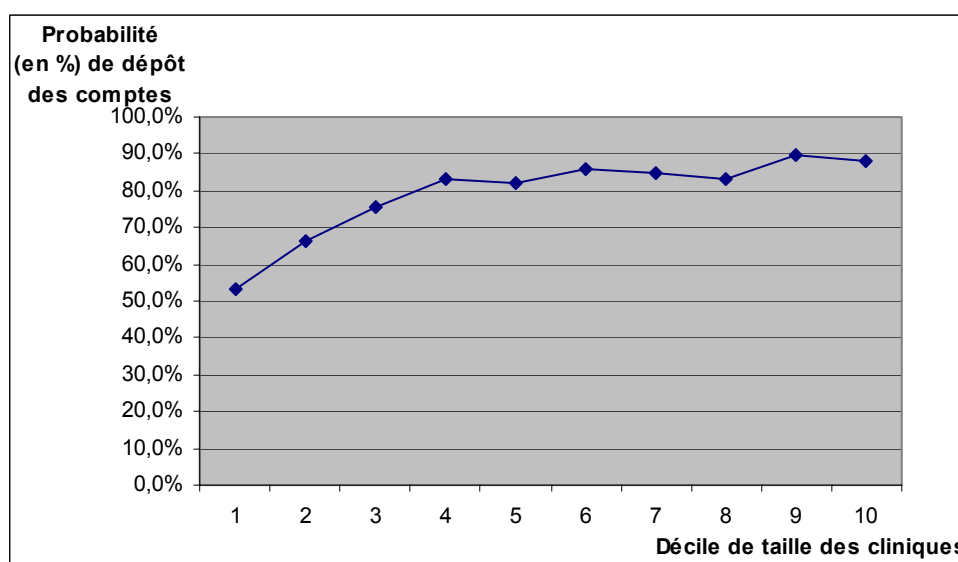
Dans les données de Diane, on ne sait pas quelle part du chiffre d'affaires déclaré imputer à la seule activité médicale de la clinique. Or le chiffre d'affaires de ces établissements est très supérieur à celui des établissements ayant une activité purement médicale. Certaines cliniques de soins de suite et de réadaptation, par exemple, déclarent un chiffre d'affaires de plus de 300 000 euros dans Diane (soit 3 % du CA total des cliniques en 2008). Les prendre en compte sans correction introduirait un biais dans la mesure du chiffre d'affaires total des cliniques. Les cliniques privées appartenant à ces groupes sont incluses dans le champ de l'étude ; toutefois, les données économiques des grands groupes auxquels elles appartiennent sont considérées comme de la non-réponse afin de ne pas surestimer les résultats.

## I. Qui sont les cliniques non répondantes ?

Dans cette partie, nous étudions si les cliniques qui déposent le moins leurs comptes ont des caractéristiques spécifiques.

La probabilité qu'une clinique dépose ses comptes est modélisée en fonction de sa taille. La taille des cliniques est approchée par les effectifs des salariés non médicaux à temps plein exerçant dans l'établissement ; elle est ici découpée en 10 classes croissantes comportant toutes le même nombre d'établissements. Plus les cliniques sont de grande taille, plus la probabilité qu'elles déposent leurs comptes est élevée (graphique 1). En outre, comme le chiffre d'affaires est croissant avec la taille des établissements (voir partie 2), plus les cliniques ont un chiffre d'affaires important, plus elles déposent leurs comptes. Par conséquent, les cliniques de petite taille, qui correspondent aux chiffres d'affaires les plus faibles, sont majoritaires parmi les cliniques non répondantes.

**Graphique 1 - Probabilité (en %) qu'une clinique a de déposer ses comptes selon sa taille (approchée par les effectifs des salariés non médicaux à temps plein) en 2008**



Note de lecture : Une clinique appartenant au 1<sup>er</sup> décile de taille (c'est-à-dire aux 10 % des cliniques ayant les effectifs des salariés non médicaux à temps plein les plus faibles) a une probabilité de déposer ses comptes de 53,3 % tandis que celle appartenant au 10<sup>ème</sup> décile a 87,8 % de chance de déposer ses comptes.

Sources : SAE, Diane.

Champ : Cliniques ayant déposé leur compte.

Dans l'étude, les cliniques sont classées en trois tailles : les petites cliniques comportent moins de 50 salariés non médicaux à temps plein, les moyennes en comprennent entre 50 et 150 et les grandes en comptent plus de 150<sup>3</sup>. La même modélisation que précédemment est réalisée avec ce découpage. Il en résulte qu'en 2008, les petites cliniques ont une probabilité moyenne de déposer leurs comptes de 69 % tandis que celle des cliniques de taille moyenne et de grande taille est plus élevée, respectivement 85 % et 87 % (tableau 1) : la non-réponse est donc nettement plus élevée chez les cliniques de petite taille. Notons, de plus, que les résultats sont proches d'une année sur l'autre.

**Tableau 1 - Probabilité (en %) de dépôt de compte d'une clinique selon sa taille**

Taille	2003	2004	2005	2006	2007	2008
petite	67	72	76	69	75	69
moyenne	79	84	87	88	90	85
grande	89	91	93	89	92	87

Note de lecture : Les petites cliniques comportent moins de 50 salariés non médicaux à temps plein, les moyennes en comprennent entre 50 et 150 et les grandes en comptent plus de 150.

Par ailleurs, les cliniques non répondantes sont surreprésentées dans les DOM-TOM (tableau 2). En effet, aucune des 35 cliniques situées dans les DOM-TOM n'a déposé ses comptes en 2008. 40 % d'entre elles les ont remis les années précédentes, mais avec retard : on peut donc supposer qu'une proportion équivalente de cliniques des DOM-TOM déposera ses comptes 2008 dans les mois à venir. Dans les autres grandes zones d'étude et d'aménagement du territoire (ZEAT), plus de 80 % des cliniques déposent leurs comptes. Toutefois, il peut arriver que certains établissements les remettent tardivement : ainsi, seulement la moitié des cliniques du Nord-est l'ont fait dans les délais impartis en 2008.

**Tableau 2 - Probabilité (en %) de dépôt de compte d'une clinique selon sa ZEAT**

ZEAT	2003	2004	2005	2006	2007	2008
DOM-TOM	33	48	44	39	44	0
Ile-de-France	79	82	87	83	87	83
Bassin parisien	77	76	80	72	77	77
Nord	79	84	81	91	88	89
Nord-est	85	87	87	77	77	47
Ouest	78	79	85	81	90	87
Sud-ouest	78	84	86	86	87	84
Centre-est	76	87	91	88	89	90
Sud-est	75	81	87	85	89	82
France entière	76	81	84	82	85	79

Enfin, la probabilité de dépôt des comptes pour une clinique s'établit à 80 % au moins pour toutes les disciplines et est stable dans le temps (tableau 3). Seules les cliniques de type « divers », qui forment un ensemble très hétérogène sans cohérence réelle, déposent deux fois moins leurs comptes que les autres.

<sup>3</sup> Les petites cliniques recouvrent les déciles allant de la taille 1 à 4, les cliniques de taille moyenne les déciles allant de 4 à 8 et les grandes cliniques les déciles supérieurs ou égaux à 8. Ces classes ont été choisies de façon à ce que les petits établissements représentent 10% du chiffre d'affaires de l'ensemble des cliniques, les moyens en représentent 30 % et les grands 60 %.

**Tableau 3 - Probabilité (en %) de dépôt de compte d'une clinique selon sa discipline**

Discipline	2003	2004	2005	2006	2007	2008
Polyclinique	82	87	88	86	89	83
Chirurgie	75	82	87	84	82	82
Dialyse	83	88	85	83	81	81
Soins de suite et de réadaptation	81	82	88	85	84	84
Psychiatrie	86	89	91	86	79	79
Divers	29	34	35	34	40	40
Ensemble	76	81	84	82	85	79

La probabilité de dépôt de compte des cliniques varie ainsi en fonction des caractéristiques des établissements, telle que leur taille. Comme la non-réponse apporte un biais aux résultats, elle nuit donc à leur qualité. Le redressement de la non-réponse permet toutefois de compenser cette perte de qualité.

## II. Le redressement de la non-réponse

### 1. Aspects théoriques du calage sur marges

Olivier Sautory, 1993, « La macro CALMAR Redressement d'un échantillon par calage sur marges », Document n° F9310 25, novembre, p 6 à 10.

#### a) Le problème

On considère une population  $U = \{1 \dots k \dots N\}$  de  $N$  individus, dans laquelle on a tiré un échantillon  $s$  de taille  $n$ . Pour tout individu  $k$  de  $U$ , on note  $\pi_k$  sa probabilité d'inclusion dans  $s$  (elle vaut  $n/N$  pour tout  $k$  dans le cas d'un sondage aléatoire simple).

Soit  $Y$  une variable d'intérêt, pour laquelle on désire estimer le total sur la population :  $Y = \sum_{k \in U} y_k$ .

L'estimateur de  $Y$  utilisé classiquement est l'estimateur de Horvitz-Thompson :

$$\hat{Y}_\pi = \sum_{k \in s} \frac{1}{\pi_k} y_k = \sum_{k \in s} d_k y_k.$$

Utiliser cet estimateur sans biais de  $Y$  revient à affecter à chaque individu de l'échantillon un poids  $d_k$  égal à l'inverse de sa probabilité d'inclusion (c'est le « coefficient d'extrapolation »  $N/n$  dans le cas d'un sondage aléatoire simple).

#### Information auxiliaire

Soit  $X_1 \dots X_j \dots X_J$   $J$  variables auxiliaires connues sur l'échantillon  $s$ , et dont **on connaît les totaux sur la population** :

$$X_j = \sum_{k \in U} x_{jk}.$$

Pour tenir compte de cette information, on va chercher à estimer le total  $Y$  de  $Y$  à l'aide d'un estimateur de forme :

$$\hat{Y}_w = \sum_{k \in s} w_k y_k,$$

où les poids  $w_k$  affectés aux individus sont « proches » (dans un sens à préciser) des poids de sondage  $d_k$ , et vérifient les **équations de calage** :

$$\boxed{\forall j = 1 \dots J, \sum_{k \in s} w_k x_{jk} = X_j}$$

On cherche donc un estimateur « peu différent » de l'estimateur de Horvitz-Thompson qui « cale » l'échantillon sur les totaux des variables auxiliaires.

#### b) Résolution théorique

On choisit une « fonction de distance »  $G$ , d'argument  $r = w_k / d_k$ , pour mesurer les distances entre les  $w_k$  et les  $d_k$  ;  $G$  doit vérifier les conditions suivantes : elle est positive et convexe, et  $G(1)=G'(1)=0$ .

Une fois la fonction G choisie, le problème consiste à déterminer les poids  $w_k$  (kCs) solutions du programme suivant (en notant les vecteurs  $x'_k = (x_{1k} \dots x_{jk})$  et  $X' = (X'_1 \dots X'_j)$ ) :

$$\text{Min}_{w_k} \sum_{k \in S} d_k G(w_k / d_k) \text{ sous la contrainte } \sum_{k \in S} w_k x_k = X$$

i.e. on minimise une somme pondérée (par les  $d_k$ ) des « distances » entre les poids de sondage  $d_k$  et les pondérations cherchées  $w_k$ , sous les contraintes du calage.

On résout ce problème en introduisant un vecteur de multiplicateurs de Lagrange  $\lambda' = (\lambda_1 \dots \lambda_j)$  ; le Lagrangien vaut :

$$L = \sum_{k \in S} d_k G(w_k / d_k) - \lambda' \left( \sum_{k \in S} w_k x_k - X \right)$$

Les conditions du 1<sup>er</sup> ordre conduisent à :

$$w_k = d_k F(x'_k \lambda)$$

où F est la fonction réciproque de la dérivée de la fonction G.

Le vecteur  $\lambda$  est déterminé par la résolution du système non linéaire de J équations à J inconnues résultant des équations de calage :

$$\sum_{k \in S} d_k F(x'_k \lambda) x_k = X \quad (E)$$

On peut résoudre numériquement ce système par la méthode itérative de Newton ; on calcule une suite de vecteurs  $\lambda^{(i)}$  définis par une relation de récurrence, en initialisant l'algorithme avec le vecteur  $\lambda^{(0)} = 0$ . La convergence est obtenue lorsque les rapports de poids  $w_k/d_k$  obtenus lors de deux itérations successives « ne bougent presque plus » :

$$\text{Max}_{k \in S} \left| \frac{w_k^{(i+1)}}{d_k} - \frac{w_k^{(i)}}{d_k} \right| < \varepsilon$$

### c) Les fonctions G usuelles

On indique pour chacune des 4 méthodes usuelles la fonction G(r) (où  $r = w_k/d_k$ ) et la fonction F(u) (où  $u = x'_k \lambda$ ).

#### 1. Méthode « linéaire »

$$-G(r) = \frac{1}{2} (r - 1)^2, r \in \mathfrak{R} \text{ et } F(u) = 1 + u \quad (u \in \mathfrak{R}).$$

La forme linéaire de F donne son nom à la méthode, dont on peut montrer qu'elle équivaut à une méthode classique d'estimation utilisant de l'information auxiliaire, appelée **estimation par régression**.

#### 2. Méthode « raking ratio »

$$-G(r) = r \text{ Log } r - r + 1, r > 0 \text{ et } F(u) = \exp u \quad (u > 0)$$

Lorsque les variables auxiliaires sont des variables catégorielles pour lesquelles on connaît les effectifs des modalités dans la population, le choix de cette fonction G conduit à une méthode classique de redressement, proposée par Deming et Stéphan, sous le nom de raking ratio ; elle est aussi connue (dans SAS en particulier) sous le nom I.P.F. (« Iterative Proportional Fitting »).

#### 3. Méthode « logit »

$$-G(r) = \left[ (r - L) \text{Log} \frac{r - L}{1 - L} + (U - r) \text{Log} \frac{U - r}{U - 1} \right] \frac{1}{A}, \text{ si } L < r < U \text{ (et } +\infty \text{ sinon)}$$

$$\text{avec } A = \left[ \frac{U - L}{(1 - L)(U - 1)} \right]$$

$$F(u) = \frac{L(U - 1) + U(1 - L) \exp(Au)}{U - 1 + (1 - L) \exp(Au)} \in ]L, U[$$

La forme logistique de la fonction F donne son nom à cette méthode, que l'on peut aussi caractériser comme étant une méthode « raking ratio » tronquée aux deux extrémités, de façon que les rapports  $w_k/d_k$  soient bornés inférieurement par L et supérieurement par U.

#### 4. Méthode « linéaire tronquée »

$$-G(r) = \frac{1}{2}(r-1)^2, \text{ si } L \leq r \leq U \text{ (+ } \infty \text{ sinon)}$$

$$F(u) = 1 + u \in [L, U]$$

#### d) Le choix de la méthode

Les principales caractéristiques des différentes méthodes sont les suivantes :

- la méthode **linéaire** est la plus rapide car elle converge toujours après deux itérations ; elle peut conduire à des poids  $w_k$  négatifs, ce qui en général ne satisfait pas le responsable d'enquête... Enfin, les poids ne sont pas bornés supérieurement, et les rapports de poids  $w_k/d_k$  peuvent prendre des valeurs que le statisticien jugera élevées (par exemple > 4).

- la méthode **raking ratio** conduit à des poids toujours positifs, mais non bornés supérieurement, d'ailleurs en général supérieurs (pour les poids les plus élevés) à ceux de la méthode « linéaire ».

- les méthodes **logit** et **linéaire tronquée** présentent l'avantage de pouvoir définir une borne inférieure L et une borne supérieure U aux rapports  $w_k/d_k$ . Toutefois, on ne peut pas choisir a priori n'importe quelles valeurs pour L et U : il existe pour L une valeur maximale  $L_{\max}$  (inférieure à 1), et pour U une valeur minimale  $U_{\min}$  (supérieure à 1). Ces valeurs dépendent des données et des marges du calage : plus la structure de l'échantillon est différente de celle de la population, plus ces valeurs sont éloignées de 1.

Dans la pratique, la détermination de ces valeurs  $L_{\max}$  et  $U_{\min}$  se fait par « approximations successives » : on fait tourner la procédure de redressement en augmentant progressivement L (valeurs inférieures à 1), et en diminuant progressivement U (valeurs supérieures à 1)... jusqu'à ce que le programme manifeste qu'il n'existe pas de solution.

Face à différents systèmes de pondération possible (on peut en obtenir théoriquement une infinité en faisant varier L et U) qui, on peut le rappeler, satisfont tous aux contraintes de calage, le responsable d'enquête doit en choisir un, et un seul. Des critères pouvant présider au choix de la pondération qui sera finalement utilisée sont les suivants :

- la plus faible dispersion ;
- la plus faible étendue ;
- l'allure générale de la distribution.

On peut ainsi souhaiter utiliser une méthode bornée (M= 3 ou 4) sans trop déformer la distribution des rapports de poids obtenus par la méthode du raking ratio par exemple : l'utilisation des valeurs  $L_{\max}$  et  $U_{\min}$  conduit en général à une très forte concentration des rapports de poids au voisinage de ces valeurs limites.

Le choix de la méthode ne peut reposer sur un critère de précision des estimateurs, car les méthodes sont toutes équivalentes (asymptotiquement). C'est à un concept, non formalisé, de « robustesse » que le statisticien fait appel, et le critère qui préside au choix est donc d'une certaine façon affaire de point de vue.

## 2. Le calage sur marges utilisé

### a) Le choix des variables auxiliaires

Le redressement des données des cliniques est réalisé à l'aide d'un calage sur marges : les établissements ayant déposés leurs comptes (qui constituent l'échantillon) deviendront alors représentatifs de l'ensemble des cliniques privées à but lucratif. Connues pour toutes les cliniques présentes dans la SAE, les variables auxiliaires sont les informations utilisées dans le calage qui, bien corrélées avec les données économiques de l'échantillon, permettent d'améliorer la précision des estimations.



Les données économiques et financières dépendent fortement de la taille des établissements, ici approchée par les effectifs des salariés non médicaux à temps plein<sup>4</sup>, ainsi que leur discipline (polyclinique, chirurgie, etc.). Comme l'échantillon est représentatif en termes de discipline et de taille, il le sera aussi en termes de données économiques et financières et les estimations seront donc de bien meilleure qualité que si elles n'avaient pas été redressées. La variable « effectifs des salariés non médicaux à temps plein » est disponible pour tous les établissements de la SAE (contrairement aux lits et places, entrées et venues ou aux effectifs des salariés médicaux) et représente bien la taille des cliniques : c'est donc une bonne variable auxiliaire pour le calage. Les marges utilisées dans le calage et la proportion de répondants pour chacune d'entre elles sont présentées dans le tableau 4.

**Tableaux 4 - Marges utilisées dans le calage**

- le nombre de cliniques par discipline

Variable	Effectifs dans la population	Pourcentage de répondants
Chirurgie	146	82%
Dialyse avec moins de 21 ETP non médicaux	26	81%
Dialyse avec 21 ETP non médicaux ou plus	26	81%
Divers	78	40%
Médecine ou obstétrique	35	66%
Polyclinique avec moins 180 ETP non médicaux	182	80%
Polyclinique avec 180 ETP non médicaux ou plus	182	88%
Psychiatrie	141	78%
Soins de suite et de réadaptation avec moins 49 ETP non médicaux	125	81%
Soins de suite et de réadaptation avec 49 ETP non médicaux ou plus	128	85%

Note : Les disciplines qui ont des effectifs de salariés non médicaux à temps plein (ETP) dispersés, sont séparées en deux classes. Elles constituent donc deux marges différentes (comprenant chacune le même nombre de cliniques), ce qui permet de prendre en compte les effets différenciés selon la taille. Malgré le petit nombre d'établissements de dialyse, ils ont été séparés en 2 catégories car les données financières de ces cliniques sont très hétérogènes.

- les effectifs des salariés non médicaux à temps plein

Variable	Effectifs dans la population	Pourcentage de répondants
Chirurgie	12 348	84%
Dialyse avec moins de 21 ETP non médicaux	289	81%
Dialyse avec 21 ETP non médicaux ou plus	1 275	86%
Divers	2 223	48%
Médecine ou obstétrique	3 034	61%
Polyclinique avec moins 180 ETP non médicaux	20 047	82%
Polyclinique avec 180 ETP non médicaux ou plus	60 225	89%
Psychiatrie	8 009	81%
Soins de suite et de réadaptation avec moins 49 ETP non médicaux	4 158	83%
Soins de suite et de réadaptation avec 49 ETP non médicaux ou plus	13 873	87%

Note : Les disciplines qui ont des effectifs de salariés non médicaux à temps plein (ETP) dispersés, sont séparées en deux classes. Elles constituent donc deux marges différentes (comprenant chacune le même nombre de cliniques), ce qui permet de prendre en compte les effets différenciés selon la taille. Malgré le petit nombre d'établissements de dialyse, ils ont été séparés en 2 catégories car les données financières de ces cliniques sont très hétérogènes.

Pour chaque discipline, un modèle linéaire simple dans lequel le chiffre d'affaires est expliqué par les effectifs des salariés non médicaux à temps plein est testé (tableau 5). Le pouvoir explicatif des modèles est fort puisque le coefficient de corrélation linéaire  $R^2$  est élevé pour chacun d'entre eux, mais il varie d'une discipline à une autre. Ainsi, le  $R^2$  est égal à 98 % pour les cliniques de soins de suite et de réadaptation, par exemple, tandis qu'il est un peu plus faible pour les « établissements divers » dont l'ensemble forme un groupe hétérogène. La

<sup>4</sup> Les lits et places et les entrées et venues sont, eux aussi, bien corrélés avec les données économiques des cliniques. Toutefois, ils n'ont pas été inclus dans les variables auxiliaires du calage car ils ne sont pas renseignés pour l'ensemble des établissements.

penne de la droite du modèle diffère aussi selon les disciplines, ce qui signifie que la corrélation entre les effectifs des salariés non médicaux à temps plein et le chiffre d'affaires des cliniques est plus ou moins forte selon les disciplines. Comme la corrélation entre le chiffre d'affaires et le nombre de salariés non médicaux à temps plein des cliniques varie en fonction des disciplines, la variable auxiliaire de taille est découpée par discipline afin de prendre en compte leur spécificité. De ce fait, la taille par discipline et le nombre d'établissements par discipline sont choisis comme variables auxiliaires du calage.

**Tableau 5 - Coefficient de corrélation linéaire R<sup>2</sup> et pente de l'équation par discipline i :**  
**Chiffre d'affaires(i) =  $\alpha$  +  $\beta$ effectifs des salariés non médicaux à temps plein(i) +  $\epsilon$**

Discipline	2008		2007	
	R <sup>2</sup> (en %)	Pente	R <sup>2</sup> (en %)	Pente
Chirurgie	83,2	83 744	85,4	83 901
Dialyse	91,3	158 838	91,9	159 295
Divers	66,8	86 001	60,5	85 213
Médecine	68,3	90 564	70,9	91 118
Obstétrique	99,0	108 498	74,0	70 619
Polyclinique	91,0	92 520	86,3	87 153
Psychiatrie	82,0	70 389	84,7	70 590
Soins de suite et de réadaptation	98,0	108 964	97,0	111 299

### ***b) La méthode de calage retenue***

À l'aide des variables auxiliaires, de nouveaux poids de sondage sont affectés aux cliniques de l'échantillon de façon à ce que la structure par taille et par discipline de l'échantillon coïncide avec celle de la population. Plusieurs méthodes de calage sont programmées dans la macro SAS CALMAR (linéaire, raking ratio, logit, linéaire tronquée et hyperbolique) ; elles se différencient par le type de fonction de distance retenu pour minimiser l'écart entre les anciennes et les nouvelles pondérations. Les différentes méthodes ont été testées pour chacune des années étudiées (tableau 7). Les méthodes linéaire tronquée et logit [0,7 ; 3,1] permettent d'obtenir la plus faible dispersion des rapports de poids (moyenne et médiane proches de 1), une variance et une étendue de ces rapports faibles ainsi qu'une distribution proche de la loi normale. C'est la méthode linéaire tronquée [0,7 ; 3,1] qui est retenue pour les calages de toutes les années étudiées car elle présente une étendue des rapports de poids plus faible pour la plupart d'entre elles, excepté en 2008. Les rapports de poids obtenus avec le calage s'étendent donc entre 0,7 et 3,1 pour l'année 2008.

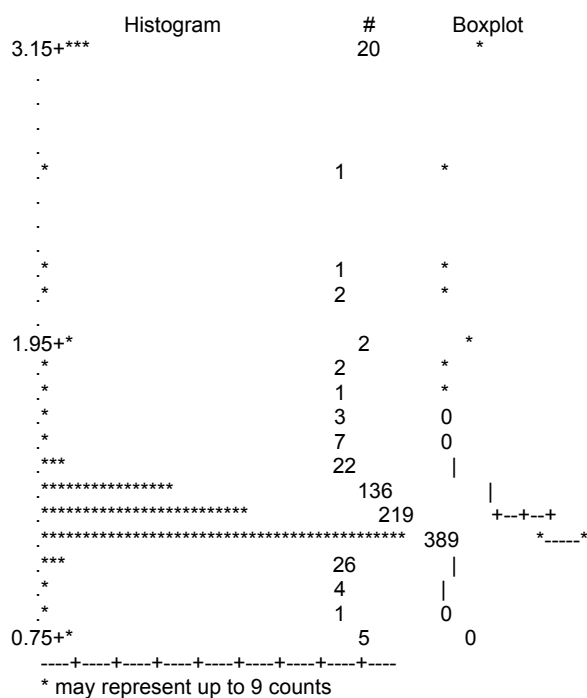
**Tableau 7 - Statistiques sur les rapports de poids obtenues avec les différentes méthodes du calage pour les données 2008**

Méthode utilisée	Moyenne	Médiane	Variance	Minimum	Maximum	Étendue
Linéaire	Éliminée car un poids est négatif					
Raking ratio	1,2711	1,1971	0,0952	0,68	4,17	3,48
Logit [0,5-4]	1,2711	1,1951	0,0946	0,60	3,60	3,00
Logit [0,7-3,1]	1,2711	1,1924	0,1010	0,70	3,10	2,40
Linéaire tronquée [0,5-4]	1,2711	1,1981	0,0942	0,50	3,78	3,28
<b>Linéaire tronquée [0,7-3,1]</b>	<b>1,2711</b>	<b>1,1974</b>	<b>0,1008</b>	<b>0,70</b>	<b>3,10</b>	<b>2,40</b>
Sinus hyperbolique (1)	1,2711	1,1970	0,0956	0,33	3,37	3,04
Sinus hyperbolique (2)	1,2711	1,2022	0,0991	0,32	3,21	2,89

## Méthode retenue

La méthode linéaire tronquée [0,7 ; 3,1] permet d'obtenir une courbe des rapports de poids unimodale, dont la distribution est centrée autour de 1,15 (graphique 3) et assez proche de celle de la loi normale.

**Graphique 2 - Rapports de poids de la méthode linéaire tronquée [0,7 ; 3,1] en 2008**



## III. Résultats du calage

Le calage sur marge modifie les poids des cliniques répondantes de façon à les rendre représentatives de l'ensemble des établissements.

Pour toutes les disciplines (excepté pour les établissements « divers »), la moyenne des rapports de poids s'établit à 1,2 ou 1,3 : ainsi, après redressement des données, une clinique de l'échantillon représente un petit peu plus qu'elle-même en moyenne (tableau 8).

**Tableau 8 - Moyenne des rapports de poids (=poids final/poids initial) obtenue après le calage selon la discipline des cliniques en 2008**

Discipline	Moyenne des rapports de poids
Chirurgie	1,2
Dialyse	1,2
Divers	2,5
Polycliniques	1,2
Psychiatrie	1,3
Soins de suite et de réadaptation	1,2

Comme les cliniques de plus petite taille sont celles qui déposent le moins leurs comptes, leur rapport de poids est plus élevé que celui des grands établissements (tableaux 9). Par conséquent, plus une clinique est petite, plus le calage lui donne un poids important pour compenser la non-réponse plus fréquente pour ce type d'établissement.

**Tableaux 9 - Moyenne des rapports de poids (=poids final/poids initial) obtenue après le calage selon la taille des cliniques en 2008**

Taille	Moyenne des rapports de poids
Petite	1,4
Moyenne	1,2
Grande	1,1

Note de lecture : Les petites cliniques comportent moins de 50 salariés non médicaux à temps plein, les moyennes en comprennent entre 50 et 150 et les grandes en comptent plus de 150.

Décile de taille	Moyenne des rapports de poids
1	1,6
2	1,5
3	1,3
4	1,3
5	1,2
6	1,2
7	1,2
8	1,2
9	1,1
10	1,1

#### IV. Quelques résultats de l'étude

En 2008, le chiffre d'affaires total des cliniques privées à but lucratif s'élève à 11,6 milliards d'euros (tableau 10 a) ; il s'établit à 9,8 milliards d'euros si l'on inclut uniquement les cliniques ayant déposé leurs comptes en 2008 (tableau 10 b). Ainsi, le calage sur marges permet de prendre en compte 15 % du chiffre d'affaires qui ne peuvent être inclus si l'on ne s'intéresse qu'aux établissements ayant déposé leurs comptes.

Notons, par ailleurs, que le redressement de la non-réponse permet de prendre en compte les fermetures, créations ou restructurations de cliniques.

**Tableaux 10 - Chiffre d'affaires total des cliniques privées entre 2003 et 2008**

**a) Chiffre d'affaires avec redressement de la non-réponse**

		CA en 2003	CA en 2004	CA en 2005	CA en 2006	CA en 2007	CA en 2008
<b>TOTAL</b>		<b>8 962</b>	<b>9 537</b>	<b>10 284</b>	<b>10 721</b>	<b>11 233</b>	<b>11 631</b>
<b>MCO</b>	Polycliniques	5 885	6 277	6 852	7 165	7 474	7 764
	Chirurgie	1 194	1 195	1 225	1 196	1 197	1 137
	Dialyse	151	182	203	234	224	241
	<b>Total MCO</b>	<b>7 230</b>	<b>7 655</b>	<b>8 280</b>	<b>8 595</b>	<b>8 896</b>	<b>9 142</b>
<b>Hors MCO</b>	Suite et réadaptation	1 065	1 194	1 278	1 343	1 509	1 639
	Psychiatrie	528	546	569	580	626	651
	Divers	139	142	158	203	202	199
	<b>Total hors MCO</b>	<b>1 732</b>	<b>1 882</b>	<b>2 005</b>	<b>2 126</b>	<b>2 337</b>	<b>2 489</b>
<b>Taille</b>	Grandes cliniques	4 572	5 175	5 608	6 010	6 478	6 850
	Moyennes cliniques	3 328	3 260	3 548	3 492	3 561	3 539
	Petites cliniques	1 062	1 103	1 129	1 219	1 193	1 243

Note : les polycliniques comprennent toutes les cliniques qui pratiquent plusieurs disciplines ainsi que celles spécialisées en médecine ou en obstétrique, très minoritaires.

La taille des cliniques est définie par le nombre d'emploi à temps plein des salariés non médicaux qui y travaillent. Les petites cliniques sont celles dont les emplois à temps pleins des salariés non médicaux sont inférieurs à 50 les moyennes cliniques celles pour lesquels ils sont compris entre 50 et 150 et les grandes cliniques celles qui en comportent plus de 150.

Sources : Greffes des tribunaux de commerce, SAE 2008, calculs DREES.

**b) Chiffre d'affaires sans redressement de la non-réponse**

		CA en 2003	CA en 2004	CA en 2005	CA en 2006	CA en 2007	CA en 2008
<b>TOTAL</b>		<b>7 577</b>	<b>8 427</b>	<b>9 315</b>	<b>9 640</b>	<b>10 182</b>	<b>9 849</b>
<b>MCO</b>	Polycliniques	4 891	5 474	6 082	6 332	6 640	6 454
	Chirurgie	935	1 024	1 090	1 049	1 080	957
	Dialyse	134	165	186	193	204	206
	<b>Total MCO</b>	<b>5 960</b>	<b>6 664</b>	<b>7 358</b>	<b>7 575</b>	<b>7 925</b>	<b>7 616</b>
<b>Hors MCO</b>	Suite et réadaptation	923	991	1 144	1 233	1 384	1 408
	Psychiatrie	461	494	527	510	558	526
	Divers	50	58	65	78	85	96
	<b>Total hors MCO</b>	<b>1 434</b>	<b>1 543</b>	<b>1 736</b>	<b>1 822</b>	<b>2 027</b>	<b>2 031</b>
<b>Taille</b>	Grandes cliniques	4 194	4 873	5 305	5 735	6 065	6 045
	Moyennes cliniques	2 636	2 731	3 116	3 021	3 193	2 900
	Petites cliniques	747	824	894	884	925	904

Note : les polycliniques comprennent toutes les cliniques qui pratiquent plusieurs disciplines ainsi que celles spécialisées en médecine ou en obstétrique, très minoritaires.

La taille des cliniques est définie par le nombre d'emploi à temps plein des salariés non médicaux qui y travaillent. Les petites cliniques sont celles dont les emplois à temps pleins des salariés non médicaux sont inférieurs à 50 les moyennes cliniques celles pour lesquels ils sont compris entre 50 et 150 et les grandes cliniques celles qui en comportent plus de 150.

Attention, le champ de ce tableau est différent de celui de l'Etudes et Résultats n°707 car on a retenu les données disponibles les plus récentes, donc des cliniques dont les comptes ont été reçus tardivement.

Sources : Greffes des tribunaux de commerce, SAE 2008, calculs DREES.

Champs : Cliniques ayant déposé leur compte.

Les écarts de résultats obtenus avec les deux méthodes sont moins importants pour les autres indicateurs, plus structurels (tableaux 11 et 12). En effet, la rentabilité économique diminue fortement entre 2007 et 2008 avec un taux s'établissant à 2,1 % du chiffre d'affaires en 2008. Cette baisse touche fortement le secteur MCO dont la rentabilité décline de 0,9 point sur la période pour s'établir à 2,1 % du chiffre d'affaires. Pour le secteur hors MCO, ce taux demeure stable à 4 % du chiffre d'affaires.

**Tableaux 11 - Rentabilité économique (en % du CA) selon la catégorie d'établissement  
entre 2003 et 2008**

**a) Chiffres obtenus avec redressement de la non-réponse**

	2003	2004	2005	2006	2007	2008
<b>Ensemble</b>	<b>1,4</b>	<b>2,1</b>	<b>3,1</b>	<b>2,8</b>	<b>2,8</b>	<b>2,1</b>
Polycliniques *	0,5	1,2	2,8	2,6	2,1	1,4
Chirurgie	1,6	2,4	2,1	2,4	3,3	1,2
Dialyse	4,1	5,8	7,2	9,1	8,2	7,6
<b>Total MCO</b>	<b>0,8</b>	<b>1,5</b>	<b>2,8</b>	<b>2,7</b>	<b>2,5</b>	<b>1,6</b>
Suite et réadaptation	3,8	5,1	5,1	3,8	4,1	4,0
Psychiatrie	5,2	4,5	4,4	5,3	5,9	5,1
Divers	-0,8	-1,7	-3,5	-6,9	-3,6	-1,5
<b>Total hors MCO</b>	<b>3,9</b>	<b>4,4</b>	<b>4,2</b>	<b>3,2</b>	<b>3,9</b>	<b>3,8</b>

\* les polycliniques comprennent toutes les cliniques qui pratiquent plusieurs disciplines ainsi que celles spécialisées en médecine ou en obstétrique, très minoritaires.

Sources : Greffes des tribunaux de commerce, SAE 2008, calculs DREES.

**b) Chiffres obtenus sans redressement de la non-réponse**

	2003	2004	2005	2006	2007	2008
<b>Ensemble</b>	<b>1,4</b>	<b>2,1</b>	<b>3,1</b>	<b>3,0</b>	<b>3,1</b>	<b>2,1</b>
Polycliniques *	0,5	1,3	2,7	2,6	2,2	1,4
Chirurgie	1,6	2,3	2,1	2,3	3,4	1,1
Dialyse	4,5	5,9	7,2	9,2	16,6	7,5
<b>Total MCO</b>	<b>0,8</b>	<b>1,5</b>	<b>2,7</b>	<b>2,7</b>	<b>2,8</b>	<b>1,6</b>
Suite et réadaptation	3,7	5,1	5,1	3,7	4,1	4,0
Psychiatrie	5,2	4,4	4,4	5,2	6,0	5,1
Divers	-0,9	-1,8	-4,7	-5,9	-3,1	-1,1
<b>Total hors MCO</b>	<b>3,9</b>	<b>4,6</b>	<b>4,5</b>	<b>3,7</b>	<b>4,3</b>	<b>4,0</b>

\* les polycliniques comprennent toutes les cliniques qui pratiquent plusieurs disciplines ainsi que celles spécialisées en médecine ou en obstétrique, très minoritaires.

Sources : Greffes des tribunaux de commerce, SAE 2008, calculs DREES.

**Tableaux 12 - Quelques indicateurs financiers selon la catégorie d'établissement  
en 2007 et 2008**

**a) Chiffres obtenus avec redressement de la non-réponse**

	2003	2004	2005	2006	2007	2008
<b>Rentabilité financière (en % du CA)</b>						
Ensemble	8,0	10,9	15,5	13,2	12,3	9,5
MCO	4,8	8,4	14,6	12,8	10,9	7,2
Hors MCO	18,6	19,0	18,6	15,0	17,7	18,5
<b>Capacité d'autofinancement (en % du CA)</b>						
Ensemble	4,8	5,2	6,3	5,2	5,4	4,4
MCO	4,3	4,8	6,2	5,0	5,1	3,9
Hors MCO	6,7	6,9	7,0	6,0	6,3	6,2
<b>Effort d'investissement (en % du CA)</b>						
Ensemble	6,8	6,4	8,4	10,5	10,1	8,6
MCO	7,2	6,6	8,2	10,9	9,4	7,5
Hors MCO	5,4	5,3	9,4	8,8	13,0	12,8

### b) Chiffres obtenus sans redressement de la non-réponse

	2003	2004	2005	2006	2007	2008
Rentabilité financière (en % du CA)						
Ensemble	8,1	10,9	15,6	13,5	13,5	9,5
MCO	4,9	8,4	14,6	12,9	12,4	7,2
Hors MCO	18,6	19,4	19,1	16,1	17,7	17,9
Capacité d'autofinancement (en % du CA)						
Ensemble	4,8	5,2	6,4	5,3	5,4	4,4
MCO	4,3	4,8	6,2	5,1	5,1	3,9
Hors MCO	6,7	6,9	7,3	6,2	6,6	6,3
Effort d'investissement (en % du CA)						
Ensemble	6,8	6,4	8,4	10,9	10,2	8,8
MCO	7,3	6,7	8,2	8,9	9,5	7,5
Hors MCO	5,1	5,1	9,3	10,5	13,3	13,7

## Conclusion

Pour conclure, les cliniques qui ne déposent pas leurs comptes ayant des caractéristiques spécifiques, on ne peut avoir qu'une vision partielle de la situation économique et financière des établissements, si l'on ne prend en compte que les cliniques répondantes. En effet, plus une clinique est grande, plus la probabilité qu'elle dépose ses comptes est élevée : ainsi, les cliniques de petite taille sont sous-représentées dans l'échantillon des cliniques répondantes, ce qui biaise les résultats. Pourtant, il est important de compenser ces pertes si l'on veut garantir la qualité des résultats.

Un calage sur marges est réalisé afin que les établissements ayant déposé leurs comptes deviennent représentatifs de l'ensemble des cliniques privées. Si l'échantillon est représentatif en termes de discipline et de taille, il les sera aussi en termes de données économiques et financières : ces variables permettent ainsi de réaliser le calage et les estimations seront donc de bien meilleure qualité que si elles n'avaient pas été redressées. Les poids des cliniques sont, par conséquent, modifiés de façon à ce que la structure de l'échantillon coïncide avec celle de l'ensemble des établissements.

Tandis que la rentabilité économique et les principaux indicateurs financiers obtenus avec le calage sont proches de ceux calculés sans redressement de la non-réponse, le chiffre d'affaires varie selon la méthode employée. Ces écarts proviennent d'une différence concernant le champ utilisé. Lorsque la non-réponse n'est pas redressée, le chiffre d'affaires est étudié sur un champ constant sur deux années car il est très sensible aux modifications de champ ; le calage, en revanche, prend en compte l'ensemble des cliniques présentes chaque année, soit un champ non constant. Ainsi, en 2008, le chiffre d'affaires s'établit à 9,6 milliards d'euros si l'on inclut uniquement les cliniques du champ constant 2007-2008 et à 11,6 milliards d'euros en redressant de la non-réponse.

En outre, le chiffre d'affaires est estimé par des strates (nombre de cliniques par discipline et effectifs de salariés non médicaux à temps complet par discipline) ayant des taux de réponse et des liens avec le chiffre d'affaires spécifiques : elles permettent donc de reconstruire un chiffre d'affaires pour l'ensemble des cliniques, le plus précisément possible. Toutefois, le nombre d'établissements étant trop faible pour certaines strates, elles ne permettent pas de donner des résultats assez précis : il est donc plus pertinent de ne présenter le chiffre d'affaires que par grandes disciplines (MCO et hors MCO), ce que nous ferons à l'avenir.