

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

SÉRIE
ÉTUDES

**DOCUMENT
DE
TRAVAIL**

La dispersion des honoraires
des omnipraticiens sur la période 1983-2004
Une application de la méthode
des régressions quantiles

Anne-Laure SAMSON

n° 62 – août 2006

MINISTÈRE DE L'EMPLOI, DE LA COHÉSION SOCIAL ET DU LOGEMENT
MINISTÈRE DE LA SANTÉ ET DES SOLIDARITÉS

Sommaire

I. Introduction	5
II. L'évolution des honoraires réels des omnipraticiens et de leur dispersion entre 1983 et 2004	6
1. Une croissance significative des honoraires réels entre 1983 et 2004.....	6
2. ... assortie d'une diminution de leur dispersion.....	10
A. Les instruments de mesure des dispersions au sein d'une population	10
B. La diminution de la dispersion des honoraires sur la période 1983-2004.....	11
3. Un lien étroit entre évolution des honoraires réels et évolution des dispersions.....	11
A. La croissance des honoraires réels et la réduction des dispersions sont concentrées dans le bas de la distribution	11
B. Une dispersion qui se réduit plus vite lorsque les honoraires augmentent rapidement.....	13
III. Les éléments explicatifs à la baisse des dispersions	15
1. Facteurs explicatifs étudiés et démarche utilisée	15
2. La féminisation des omnipraticiens et la progression de l'activité féminine	17
A. Un point sur la structure des omnipraticiens par sexe	17
B. Une homogénéisation des honoraires féminins.....	19
C. Une réduction des disparités d'honoraires entre les hommes et les femmes	20
3. La progression de l'activité des omnipraticiens en début de carrière.....	22
A. Un point sur la structure des omnipraticiens par âge	22
B. Une homogénéisation des honoraires des omnipraticiens en début de carrière	23
C. Une réduction des disparités d'honoraires entre les omnipraticiens de moins et de plus de sept ans d'ancienneté	25
4. L'évolution de la densité médicale dans le département ou la région d'exercice de l'omnipraticien	27
A. Un point sur l'évolution de la densité médicale entre 1983-2004	27
B. Les conséquences en terme de dispersion des honoraires.....	30
C. Une réduction des disparités d'honoraires entre praticiens exerçant dans des régions inégalement peuplées en omnipraticiens	32
IV. Quelques pistes à explorer	33
V. Annexe 1 : Résultats complémentaires sur les régressions quantiles.....	35
1. Graphiques des coefficients estimés et intervalles de confiance associés.....	35
2. Les résultats des régressions quantiles effectuées par sexe.....	39
3. Le calcul des rendements de l'ancienneté dans l'exercice de la médecine :	40
4. Les résultats des régressions interquantiles.....	42
VI. Annexe 2 : La méthode des régressions quantiles.....	45
1. Introduction.....	45
2. Les prémices des régressions quantiles	46
A. Hogg (1975) et les droites de pourcentage	46
B. La recherche d'alternatives aux Moindres carrés ordinaires.....	47
3. Rappels sur les fonctions quantiles	49
A. Aperçu des principales notations utilisées dans les différentes sections.....	49
B. Définition du $\theta^{\text{ème}}$ quantile non conditionnel d'une variable aléatoire.....	49
C. Lien entre la fonction de répartition et la fonction quantile d'une variable aléatoire	50
D. De la fonction quantile théorique à la fonction quantile empirique.....	53

4. Le point de départ des régressions quantiles : les quantiles non conditionnels sont solution d'un programme d'optimisation	56
A. Écriture du programme d'optimisation.....	56
B. Résolution analytique du programme de minimisation dans le cas des quantiles non conditionnels théoriques	62
C. Conséquences et intuitions.....	64
5. Généralisation à l'estimation des régressions quantiles	66
A. Hypothèses du modèle.....	66
B. Programme d'optimisation	67
C. Conséquences	70
6. Propriétés asymptotiques	74
A. Propriétés asymptotiques des quantiles de régressions.....	74
B. Estimation de la matrice de variance-covariance asymptotique par Bootstrap :.....	75
C. Les tests statistiques :.....	77
7. Les propriétés des régressions quantiles	78
A. Équivariance.....	78
B. Invariance aux transformations monotones :	82
8. Les avantages des régressions quantiles par rapport aux régressions par moindres carrés ordinaires	82
A. Une méthode semi-paramétrique.....	82
B. Une estimation de la distribution globale de Y :	83
C. La robustesse aux « outliers » (valeurs atypiques) :	83
D. Les régressions quantiles plutôt que les moindres carrés ordinaires ou les régressions robustes	88
E. Les régressions quantiles permettent de détecter l'hétéroscédasticité des données	90
9. Mise en œuvre des régressions quantiles grâce au logiciel STATA	93
A. Les commandes utilisées pour la régression médiane	94
B. Les commandes utilisées pour l'ensemble des régressions quantiles	94
C. Interprétation des résultats obtenus sous STATA pour le cas particulier de la régression médiane	95
VII. Bibliographie	97
1. Étude de la dispersion des honoraires des omnipraticiens :	97
2. La méthode des régressions quantiles :	98

I. Introduction

En se concentrant sur la population des omnipraticiens, cette étude s'attache à décrire le plus précisément possible l'évolution de leurs honoraires entre 1983 et 2004, mais surtout à en mesurer et à en expliquer la dispersion.

De nombreuses études déjà réalisées sur les honoraires des médecins, et en particulier sur ceux des omnipraticiens, décrivent le niveau et l'évolution des honoraires moyens. À l'inverse, l'objectif est ici de se concentrer sur la dispersion par rapport à cette moyenne et entre les praticiens (par exemple entre des omnipraticiens en début de carrière et installés depuis plusieurs années ou entre les hommes et les femmes). C'est le meilleur moyen de juger de la répartition des richesses au sein de cette population.

C'est le panel de médecins libéraux issu du Système national interrégimes (SNIR) géré par la Caisse nationale d'assurance maladie des travailleurs salariés (encadré 1) qui sert de support à cette étude.

Dans un premier temps, il s'agira de faire le point sur l'évolution des honoraires réels et de leur dispersion entre 1983 et 2004, en montrant que la baisse de la dispersion observée globalement est due au rattrapage du bas de la distribution, composé principalement de femmes et de praticiens en début de carrière.

Dans un second temps, l'objectif sera d'expliquer de quelle manière cette homogénéisation des honoraires s'est réalisée, en analysant principalement trois facteurs explicatifs. Pour cela, des indicateurs courants de dispersion ainsi qu'une méthode économétrique, les régressions quantiles, seront mobilisées.

Les annexes de ce document permettent de mieux comprendre les objectifs et la mise en œuvre de cette méthode, encore peu utilisée mais qui s'avère être un outil économétrique performant pour l'analyse des dispersions et plus globalement, des inégalités.

II. L'évolution des honoraires réels des omnipraticiens et de leur dispersion entre 1983 et 2004

Il convient tout d'abord d'étudier l'évolution des honoraires réels des omnipraticiens sur la période 1983-2004 ainsi que celle de leur dispersion, dont le rythme d'évolution est très étroitement lié à la croissance des honoraires.

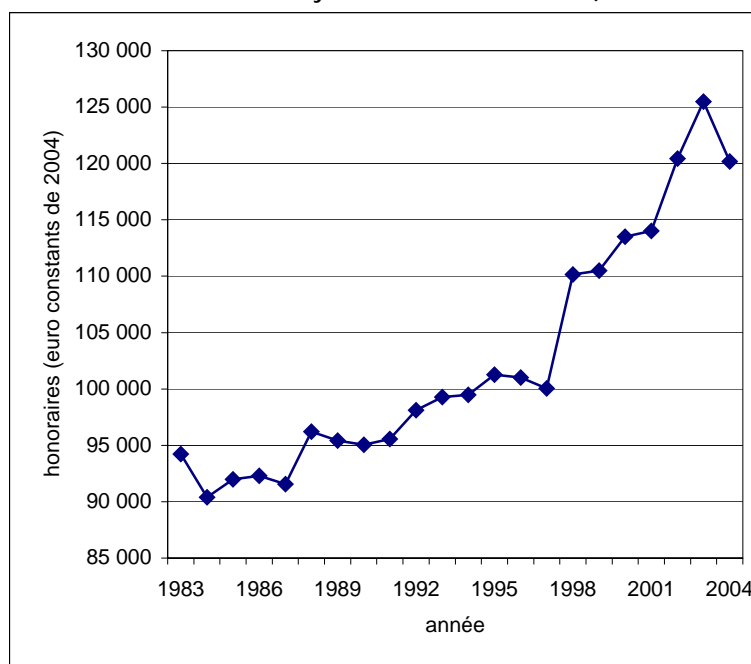
1. Une croissance significative des honoraires réels entre 1983 et 2004...

Les honoraires réels des omnipraticiens « actifs à part entière » ont augmenté de manière significative entre 1983 et 2004, leurs honoraires annuels moyens (calculés en euros constants de 2004) étant passés de 94 231 à 120 188 euros en 2004, soit une hausse de 27,5 %¹.

Sur cette période, la croissance des honoraires n'est cependant pas continue et certaines années sont marquées par une progression de plus grande ampleur (1988, 1992, 1998, 2000, 2002 et 2003). Ces quelques années sont d'autant plus remarquables qu'elles sont isolées et suivies d'une relative stabilité des honoraires (graphique 1).

L'année 2004 est atypique car elle est marquée par la forte diminution des honoraires réels par rapport aux deux années précédentes, phénomène qui n'avait plus été observé de façon si marquée depuis 1984. C'est le contrecoup de deux années de croissance exceptionnelle des honoraires : la hausse des tarifs commencée en 2002 et achevée en 2003 a fortement profité aux omnipraticiens ces années là ; en contrepartie, la stagnation des tarifs (en euros constants de 2004) n'a plus tiré les honoraires à la hausse l'année suivante.

Graphique 1 : Évolution des honoraires réels moyens entre 1983 et 2004 (en euros constants de 2004)



¹ Ces résultats ne sont pas exactement identiques à ceux obtenus grâce aux données du SNIR de la CNAMTS. Dans ces fichiers, la hausse des honoraires réels est de 24,4 % sur la même période (encadré 1).

Il est possible d'expliquer certaines années de croissance par les politiques tarifaires de la période : 1992, 1998, 2002 et 2003 sont des années de revalorisation du tarif des consultations, intervenant après plusieurs années de gel.

Alors que les revalorisations tarifaires sont annuelles, voire bi-annuelles, entre 1983 et 1987, la consultation à 100 F de 1992 est instaurée après deux années sans revalorisations, et celle à 115 F, de 1998, après trois années. Le « C à 20 € » a marqué les esprits : il est l'aboutissement des deux revalorisations tarifaires de l'année 2002 (en février et juillet) alors que les généralistes attendaient une renégociation de leurs tarifs depuis quatre ans. Cette même année est créée une consultation annuelle approfondie pour les médecins généralistes réservée aux patients souffrant d'une affection longue durée². Enfin, en septembre 2003, la mise en place du forfait pédiatrique³ pour les omnipraticiens pourrait être à l'origine d'une partie de la forte croissance des honoraires cette année-là.

La conjoncture épidémiologique est également à prendre en compte. À ce titre, l'année 1988 connaît une très forte épidémie de grippe et l'année 2000 est caractérisée par une importante épidémie de gastro-entérites.

De plus, le passage à l'an 2000 et les modifications des systèmes informatiques qui lui sont associées fin 1999, ont entraîné des retards dans la liquidation des feuilles de soins, rattrapés en 2000.

Toutefois, ces éléments explicatifs doivent être utilisés avec précaution, dans la mesure où d'autres années, également marquées par de fortes épidémies de grippe (1993, 1995) ou des revalorisations tarifaires de grande ampleur (1990, 1994, 1995), ont connu un taux de croissance des honoraires très faible, voire négatif.

² Fixée initialement à 23 € elle passe à 26 € en 2003.

³ Auparavant réservé aux pédiatres, ce forfait pédiatrique (noté FPE) permet à l'omnipraticien de majorer le prix d'une visite ou d'une consultation pour un enfant de moins de deux ans, effectuée dans les 8 jours qui suivent la naissance, au cours du 9^{ème} ou du 10^{ème} mois ou au cours du 24^{ème} ou 25^{ème} mois. Ces consultations ou visites spécifiques comportent un interrogatoire, un examen complet, des prescriptions préventives, thérapeutiques ou d'examen complémentaires. Elles doivent, entre autres, donner lieu à une mise à jour du carnet de santé de l'enfant. Le tarif de la consultation est alors fixé à 25 € au lieu des 20 € habituels.

Encadré 1 : Source et champ de l'étude

L'étude porte sur la dispersion des honoraires des omnipraticiens libéraux. Ce terme recouvre les médecins généralistes ainsi que les médecins possédant un Mode d'exercice particulier (MEP) comme l'acupuncture ou l'homéopathie.

Source des données

Le panel de médecins libéraux utilisé, fourni par la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS), reprend les données issues des statistiques (en date de remboursement) du Système national interrégimes (SNIR). En particulier, on y trouve le montant des honoraires des omnipraticiens (honoraires totaux annuels, dépassements présentés au remboursement et frais de déplacement), les caractéristiques de leur activité (activité annuelle décomposée, en distinguant notamment les consultations et les visites, année de thèse et d'installation en libéral, région et département d'exercice, secteur conventionnel, mode d'exercice, présence ou non d'un mode d'exercice particulier...) ainsi que quelques caractéristiques socio-démographiques (identifiant, année de naissance, sexe).

Couvrant la période 1979-2004, c'est un panel représentatif de la population des omnipraticiens français en exercice une année donnée, puisque tous ceux nés au mois de mai y sont présents ; il est non cylindré car complété, chaque année, par les observations relatives aux nouveaux installés.

Représentativité des honoraires moyens du panel par rapport à ceux du SNIR

La représentativité des honoraires du panel croît légèrement avec le temps et c'est à partir de l'année 1998 que les honoraires annuels moyens du panel coïncident presque exactement avec le montant publié par la CNAMTS à partir des données du SNIR ; ils étaient de 2 à 5 % inférieurs les années précédentes. Dès lors, l'évolution des honoraires réels moyens est un peu plus élevée dans le panel : +27,5 % entre 1983 et 2004, contre 24,4 % avec les données du SNIR. En revanche, entre 1998 et 2004, la progression des honoraires (en euros constants de 2004) est de 10 % environ pour les deux sources, signe de la fiabilité des résultats obtenus dans cette étude à partir des données du panel sur cette période.

Champ

Deux approches sont habituellement retenues pour analyser les honoraires des omnipraticiens :

- Soit l'étude porte sur ceux ayant perçu au moins un euro d'honoraires au cours de l'année. C'est le champ régulièrement conservé dans les publications de la DREES sur les revenus des médecins, dans un souci de cohérence avec les données fiscales utilisées.
- Soit elle se concentre sur les honoraires des « actifs à part entière ». Sont alors exclus, les omnipraticiens venant de s'installer ou cessant leur activité dans l'année, les non-conventionnés, les plus de 65 ans et les praticiens hospitaliers à temps plein. C'est l'approche généralement mise en œuvre dans les statistiques de la CNAMTS.

Ici, c'est cette deuxième approche qui a été adoptée, ce choix n'étant pas neutre pour la mesure de l'évolution des dispersions d'honoraires. En effet, conserver dans le champ d'étude les omnipraticiens qui s'installent ou qui cessent leur activité au cours de l'année ou les plus de 65 ans (au total, environ 10 % des observations de la base) sous-estime les honoraires puisque aucun indicateur ne permet de mensualiser les données. Et même si cette sous-estimation reste faible, les dispersions observées sont en revanche largement surestimées (par exemple, le rapport interquartile (interdécile) de 1,80 (3,28) en 2004 passe à 1,88 (3,71) si toutes les observations de la base sont utilisées).

Par ailleurs, pour des raisons de fiabilité de l'information, l'étude porte sur les omnipraticiens exerçant en France métropolitaine et seules les années 1983 à 2004 sont conservées.

Encadré 2 : La dispersion des revenus des omnipraticiens

Le passage des honoraires aux revenus

La DREES publie régulièrement des données chiffrées sur le niveau et l'évolution des revenus libéraux moyens des omnipraticiens (les honoraires nets des charges professionnelles), ces derniers étant calculés, pour chaque année, en appliquant aux honoraires moyens du SNIR le taux de charge moyen obtenu dans les déclarations des Bénéficiaires non commerciaux (BNC) de la Direction générale des impôts⁴.

L'objectif étant de travailler, à terme, sur la dispersion des revenus des omnipraticiens, tout en bénéficiant des données très détaillées du panel (encadré 1), les charges individuelles ont été estimées à partir des données des BNC puis appliquées aux honoraires individuels du panel d'omnipraticiens⁵. Ce travail n'étant pas terminé à la date de l'étude, seuls quelques résultats peuvent être présentés ici, à titre de comparaison.

Un aperçu de la dispersion des revenus entre 1993 et 2003⁶

- Les revenus de 2003 sont, en moyenne, de 47 % inférieurs aux honoraires (de l'ordre de 63 000 euros⁷, contre 119 000 euros pour les honoraires moyens).
- Les taux de charges sont très variables d'un omnipraticien à l'autre, car ils sont très sensibles à la commune d'installation (zone rurale ou urbaine)⁸, à la région d'exercice, au nombre d'omnipraticiens dans le cabinet (l'activité en groupe est un moyen de réduire les charges fixes de chaque omnipraticien), à l'ancienneté (les dépenses sont plus lourdes en début de carrière) et surtout au niveau d'activité et d'honoraires des omnipraticiens (les plus « actifs » étant aussi ceux dont les taux de charges sont les plus faibles, en raison de la forte proportion de charges fixes⁹). En conséquence, si en 2003, la médiane des taux de charges est de 44 % (en pourcentage des honoraires nets de débours et rétrocessions), ces taux varient entre 39 % (premier décile) et 55 % (neuvième décile).
- Il en résulte une dispersion des revenus bien supérieure à celle des honoraires. En 2003, les rapports interdéciles d'honoraires sont de 3,85, alors que ces mêmes rapports passent à 6,81 quand l'étude porte sur les revenus. Le taux de charges étant décroissant avec le niveau des honoraires, les bas revenus sont encore plus éloignés des hauts revenus que ne l'étaient les honoraires faibles des honoraires élevés.
- Toutefois, sur la période 1993-2003, la dispersion des revenus évolue de manière comparable à celle des honoraires. Comme pour les honoraires, la croissance des revenus est concentrée dans le bas de la distribution : les revenus du premier décile ont augmenté de 60 % sur la période, contre seulement 20 % pour ceux du neuvième décile. Par ailleurs, c'est la diminution de la dispersion des revenus compris entre le bas et le milieu de la distribution qui contribue le plus à la baisse des dispersions au plan global : passant de 3,6 en 1993 à 2,9 en 2003, le rapport entre la médiane et le premier décile diminue beaucoup plus fortement que le rapport entre le neuvième décile et la médiane qui, sur la même période, se réduit de 1,87 à 1,75.

⁴ Cf. Legendre, N. (2005) pour une explication détaillée de la méthode. Il serait possible, puisque les honoraires sont également présents dans les BNC, de travailler directement à partir des revenus observés dans les BNC. Néanmoins, les honoraires des BNC diffèrent fortement des honoraires fournis par le SNIR et ces derniers sont plus fiables.

⁵ Les charges individuelles sont estimées en utilisant les honoraires comme unique variable explicative (c'est la seule variable commune aux deux bases de données). Quatre méthodes différentes sont conservées afin de tester la robustesse du revenu estimé : les estimations par moindres carrés ordinaires effectuées en utilisant une fonction polynomiale d'ordre 3, puis d'ordre 2 des honoraires sont ainsi comparées à la régression médiane et à l'estimation de la relation à l'aide d'une fonction linéaire par morceaux.

⁶ Ce n'est qu'à partir de 1993 que les données individuelles des charges sont disponibles. Pour les années antérieures, la DGI ne fournit que des données fiscales agrégées.

⁷ Les chiffres obtenus sont assez proches de ceux publiés par Legendre, N. (2006) à partir des données des BNC et du SNIR, puisque les revenus des omnipraticiens sont évalués à 65 144 euros en 2003 (étant donné un taux de charge de 44% environ). La légère différence entre les deux montants moyens peut s'expliquer. Bien que les taux de charges soient calculés à partir des mêmes données (les BNC), les honoraires sur lesquels ils sont appliqués diffèrent : ceux du panel sont un peu plus faibles que les honoraires constatés dans les données du SNIR non exhaustives (à titre d'exemple, les honoraires moyens courants en 2003 sont de 121 000 euro dans les données du SNIR contre 118 000 dans le panel).

⁸ Cf. Breuil-Genier, P. (2003). Les taux de charges sont généralement croissants avec le degré d'urbanisation de la commune d'exercice du médecin : « [ils] sont particulièrement dispersés, variant de 43,7% dans les unités urbaines de moins de 5000 habitants à 47,8% dans les unités urbaines de 200 000 à 1 999 999 habitants ou les pôles urbains, et atteignant même 50% dans l'agglomération parisienne ».

⁹ Cf. Audric, S. (2003). Les charges fixes incluent le loyer, les frais de déplacement et les cotisations sociales personnelles, ces deux dernières composantes étant particulièrement fortes chez les omnipraticiens. Dès lors, « les taux de charges décroissent au fur et à mesure que les honoraires augmentent ».

2. ...assortie d'une diminution de leur dispersion

A. Les instruments de mesure des dispersions au sein d'une population

Afin de caractériser la distribution des honoraires des omnipraticiens, cette étude utilise fréquemment la notion de quantiles¹⁰. L'évolution de la dispersion des honoraires sur la période est alors quantifiée au moyen du rapport interquartile et du rapport interdécile, deux déclinaisons du rapport interquartile (tableau 1).

Ces indicateurs sont intéressants car il est possible de les décomposer afin d'analyser les dispersions à différents points de la distribution des honoraires.

Ainsi, le rapport interquartile, noté $\frac{Q3}{Q1}$ est tel que $\frac{Q3}{Q1} = \frac{Q3}{Q2} * \frac{Q2}{Q1}$.

Le premier indicateur (le rapport entre les honoraires du troisième quartile et de la médiane) caractérise les dispersions du « haut de la distribution », alors que le second (le rapport entre les honoraires de la médiane et du premier quartile) caractérise celles du « bas de la distribution ».

La même décomposition peut être effectuée pour le rapport interdécile, puisque $\frac{D9}{D1} = \frac{D9}{D5} * \frac{D5}{D1}$.

Il faut noter que les indicateurs d'inégalité étudiés dans cette partie ne caractérisent que la population des omnipraticiens dans sa globalité et ne permettent pas de juger de sa répartition au sein de sous-populations. Dans la suite de l'étude, une démarche similaire sera adoptée à des degrés plus fins : par ancienneté, sexe et densité médicale dans la région d'exercice de l'omnipraticien.

Il existe une littérature abondante sur la mesure du degré d'inégalités au sein d'une population, et d'autres indicateurs auraient pu être envisagés. Certains permettent de prendre en compte l'inégalité de l'ensemble de la répartition, et pas seulement entre les déciles extrêmes. Les courbes de Lorenz (courbes de fréquence cumulative comparant la distribution effective des honoraires à la distribution uniforme caractérisant la situation d'équirépartition) et les indices de Gini qui leur sont associés, sont, à ce titre, des instruments d'analyse courants.

Des mesures normatives de l'inégalité, qui donnent par exemple davantage de poids au bien être des plus pauvres (notamment les indices de Gini généralisé et d'Atkinson), apportent, par ailleurs, un autre éclairage.

¹⁰ Les quantiles permettent de diviser une population ordonnée en sous-groupes de taille identique. Les quartiles la partagent en quatre parties égales et les déciles en dix. Cela signifie par exemple que la rémunération d'un médecin se situant au premier décile de la distribution des honoraires est supérieure à celle des 10 % de médecins les plus pauvres, mais inférieure à celle des 90 % restant.

B. La diminution de la dispersion des honoraires sur la période 1983-2004

L'accroissement des honoraires moyens des omnipraticiens entre 1983 et 2004 s'est accompagné d'une diminution de leur dispersion.

Ainsi, les honoraires des omnipraticiens situés dans le quart supérieur étaient 2,05 fois plus élevés que ceux des omnipraticiens du quart inférieur en 1983. Ils ne l'étaient que 1,9 fois en 1992 et 1,8 fois en 2004. Les rapports interdéciles, parce qu'ils prennent en compte les honoraires extrêmes de la distribution, sont beaucoup plus élevés mais ils diminuent également plus fortement sur la période : les 10 % d'omnipraticiens les mieux rémunérés gagnaient en 1983 4,2 fois plus que les 10 % les moins bien rémunérés, contre 3,8 fois en 1992 et 3,3 fois en 2004 (tableau 1).

Tableau 1 : Évolution de la dispersion des honoraires entre 1983 et 2004

	1983	1986	1989	1992	1995	1998	2001	2004
Rapports interquartiles								
Q3/Q1	2,05	2	1,98	1,91	1,88	1,81	1,78	1,8
Q3/Q2	1,36	1,33	1,35	1,33	1,32	1,3	1,3	1,32
Q2/Q1	1,5	1,5	1,47	1,43	1,42	1,39	1,37	1,37
Rapports interdéciles								
D9/D1	4,16	4,05	4	3,78	3,58	3,37	3,32	3,28
D9/D5	1,71	1,68	1,71	1,68	1,67	1,63	1,62	1,65
D5/D1	2,44	2,4	2,34	2,25	2,14	2,07	2,04	1,99

Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004

Note de lecture :

- En 2004, D9/D1=3,28 (Q3/Q1=1,8) signifie que les 10 % (25 %) d'omnipraticiens aux honoraires les plus élevés gagnent au moins 3,28 (1,8) fois plus que les 10 % (25 %) d'omnipraticiens aux honoraires les plus faibles.
- Q2 et D5 sont deux notations qui représentent le même quantile, la médiane (Q2=D5). C'est la valeur des honoraires telle que la moitié des omnipraticiens gagne davantage tandis que l'autre moitié gagne moins.

3. Un lien étroit entre évolution des honoraires réels et évolution des dispersions

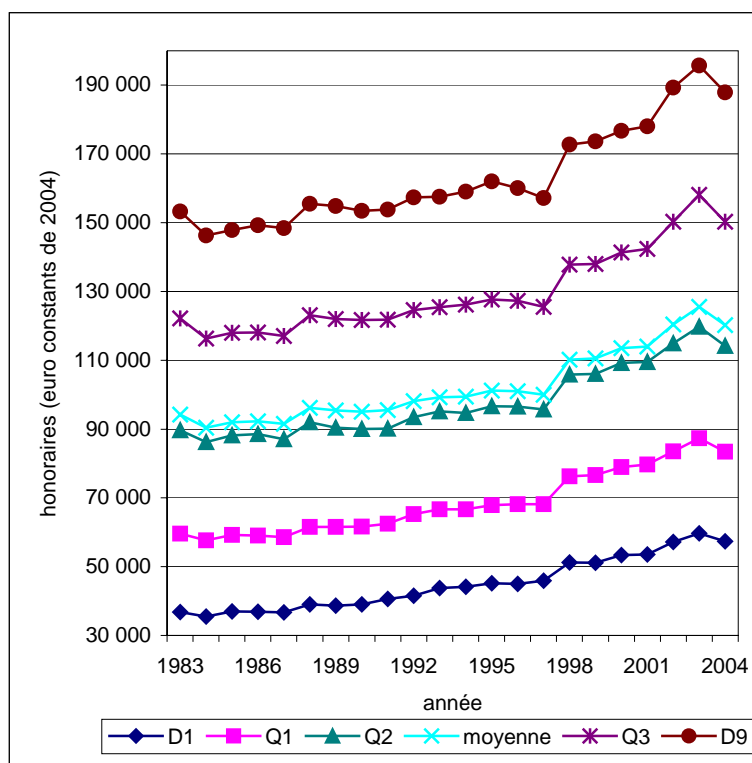
A. La croissance des honoraires réels et la réduction des dispersions sont concentrées dans le bas de la distribution

Cette homogénéisation trouve son origine dans une croissance relative des honoraires réels plus forte dans le bas de la distribution¹¹. En effet, les honoraires des omnipraticiens du premier décile¹² ont crû, en un peu plus de vingt ans, de 56 % en euros constants, passant de 36 800 euros en 1983 à 57 400 euros en 2004 et ceux des omnipraticiens du premier quartile de 40 %. En revanche, la croissance a été sensiblement moindre pour les omnipraticiens du troisième quartile et du neuvième décile, de l'ordre de 23 % dans les deux cas (graphique 2).

¹¹ Dans toute l'étude, le « bas de la distribution » se rapporte aux omnipraticiens situés en dessous du premier décile ou du premier quartile des honoraires, le « milieu de distribution » concerne ceux qui se trouvent autour de la médiane (entre le premier et le troisième quartile) et le « haut de la distribution » est utilisé pour ceux situés dans le quartile ou le décile supérieur de la répartition des honoraires.

¹² Il ne s'agit pas ici des honoraires moyens des omnipraticiens du premier décile, mais de la valeur des honoraires qui sépare les omnipraticiens du premier et du deuxième décile.

Graphique 2: Évolution de la distribution des honoraires en euros constants de 2004

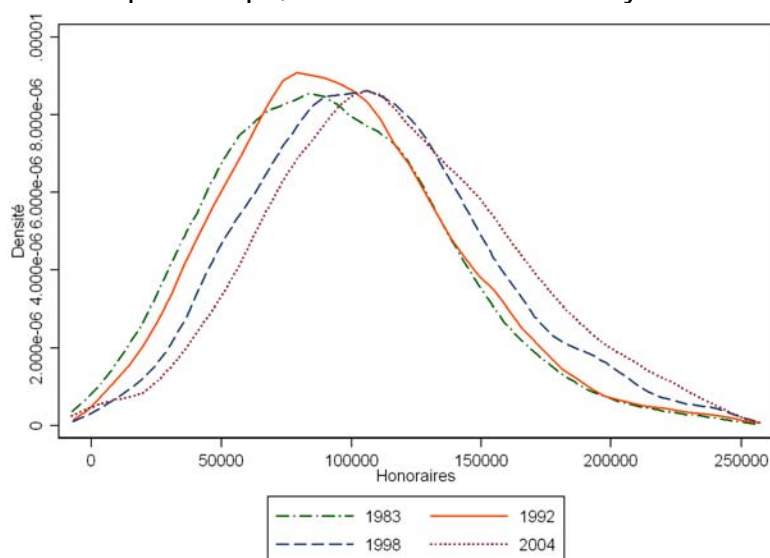


Source : Panel de médecins libéraux du SNIR (CNAMTS), 1983-2004

L'évolution de la répartition des honoraires entre 1983 et 2004 (graphique 3) illustre cette homogénéisation : c'est, de façon combinée, l'augmentation des honoraires des omnipraticiens les moins bien rémunérés, la baisse de la part des très faibles honoraires, et la hausse, relativement moins importante, des honoraires des omnipraticiens du haut de la distribution, qui ont conduit, sur la période, au déplacement vers la droite de la distribution.

C'est la dispersion des honoraires compris entre le bas et le milieu de la distribution (qui par ailleurs, contribue le plus aux dispersions globales) qui a le plus diminué sur la période. Alors que le rapport entre le troisième quartile et la médiane a oscillé entre 1,36 et 1,33 jusqu'en 1995, puis s'est stabilisé autour de 1,30 les années suivantes, le rapport entre les honoraires correspondant à la médiane et au premier quartile a diminué continûment et à un rythme plus élevé : il est passé de 1,5 en 1983 à 1,43 en 1992 et 1,37 en 2004. En 2004, les dispersions entre le haut et le bas de la distribution sont donc très proches.

Graphique 3 : Évolution de la distribution des honoraires des omnipraticiens entre 1983 et 2004 (estimation non paramétrique, basée sur la méthode des noyaux d'honoraires)



Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004 (honoraires en euros constants de 2004)

B. Une dispersion qui se réduit plus vite lorsque les honoraires augmentent rapidement

Cette réduction de la dispersion des honoraires s'est réalisée en plusieurs étapes, directement liées à leur rythme de croissance, comme en témoigne le déplacement progressif des courbes de densité entre 1983 et 2004 (graphique 3).

Entre 1983 et 1992, période de faible croissance des honoraires (+0,4 % en moyenne par an) la réduction de la dispersion résulte du relèvement des honoraires du bas de la distribution (le rapport entre le premier décile et la médiane diminue de 2,44 à 2,25), le haut de la distribution restant quasiment inchangé.

En revanche, à partir de 1992, l'ensemble de la distribution se déplace vers la droite, de façon très marquée entre 1992 et 1998 (période pendant laquelle le taux de croissance annuel moyen des honoraires est par ailleurs très élevé, de 1,7 %), puis moins fortement ensuite. Signe d'une diminution des dispersions à tous les niveaux de la distribution, le rapport entre la médiane et le premier décile se réduit de 1,43 à 1,39 et le rapport entre le troisième quartile et la médiane diminue de 1,33 à 1,30. Ce mouvement perdure entre 1998 et 2001, mais à un rythme moins soutenu, en lien direct avec une croissance des honoraires un peu plus faible, rapportée à 0,8 % en moyenne par an.

L'année 2002 est une exception. En effet, la forte progression des honoraires enregistrée cette année-là (+2,8 %) s'accompagne d'une hausse de leur dispersion dans le haut de la distribution. Or toutes les années qui avaient auparavant connu une croissance élevée des honoraires (notamment 1992, 1998 et 2000) étaient aussi celles où leur dispersion avait le plus diminué, à tous les points de la distribution.

Enfin, entre 2002 et 2004, on observe une stabilisation des honoraires réels moyens (la hausse des honoraires entre 2002 et 2003 est exactement compensée par la baisse de même ampleur entre 2003 et 2004), ainsi que de leur concentration.

Encadré 3 : Les facteurs théoriques expliquant la dispersion des honoraires entre omnipraticiens

Plusieurs éléments peuvent expliquer la dispersion des honoraires à une date donnée ainsi que son évolution au cours d'une période. Les facteurs répertoriés ci-dessous influencent surtout le niveau d'activité des omnipraticiens. Toutefois, dans la mesure où les honoraires des omnipraticiens du secteur 1 (qui représentent, selon les années, 80 à 90 % des observations de la base) sont directement liés à leur activité et que la tarification adoptée est le seul élément supplémentaire qui entre en compte pour déterminer ceux des omnipraticiens du secteur 2, il est possible de considérer l'ensemble des facteurs ci-dessous comme source de dispersions d'honoraires.

Le **volume d'offre de soins** choisi par l'omnipraticien influence la dispersion des honoraires. L'offre des hommes est, à cet égard, beaucoup plus forte que celle des femmes, ces dernières travaillant moins de jours par semaine (moins souvent que les hommes les mercredi, samedi et dimanche) et s'absentant en moyenne un jour de plus au cabinet par semaine¹³. Leurs honoraires sont en conséquence plus faibles.

Par ailleurs, si la dépense de soins, en forte augmentation sur la période, concerne en moyenne l'ensemble des omnipraticiens, **le volume de la demande de soins adressée à chacun d'entre eux** est variable et représente une source de dispersion dans le niveau de leurs honoraires :

- La demande varie en fonction de l'ancienneté de l'omnipraticien. Celle qui s'adresse aux omnipraticiens en début de carrière (qui exercent depuis moins de sept années environ¹⁴) est beaucoup plus faible que celle à laquelle les omnipraticiens déjà installés doivent répondre. En effet, les premiers, en phase de constitution de leur clientèle, ont une activité et des honoraires plus faibles que les autres, qui ont stabilisé la leur.
- Par ailleurs, la demande varie en fonction du département d'exercice du médecin. Bien qu'au niveau national la densité médicale (le nombre de médecins pour cent mille habitants) soit très élevée et en augmentation sur une grande partie de la période, il existe un fort déséquilibre entre des départements très densément peuplés en omnipraticiens dans lesquels ces derniers peuvent subir un rationnement de leur activité (notamment dans les régions PACA et Languedoc-Roussillon) et des départements faisant face à une pénurie et dans lesquels la demande de soins adressée à chaque praticien est plus forte (dans les régions Centre ou Bourgogne par exemple). Parce qu'elle influence le nombre de leurs patients potentiels, la densité médicale locale et son évolution peuvent donc être à l'origine d'une partie de la dispersion des honoraires des omnipraticiens¹⁵.

Enfin, le **choix du mode d'activité**, c'est-à-dire du secteur de conventionnement et du type de pratique (Mep ou non), peut également être à l'origine de dispersions.

- Les omnipraticiens conventionnés en secteur 1 n'ont pas le choix de la tarification adoptée (exceptés les 1 à 2 % ayant un droit permanent à dépassement) alors que ceux du secteur 2 fixent librement (avec « tact et mesure ») leurs tarifs.
- Par ailleurs, certains médecins ont acquis une spécialisation, un mode d'exercice particulier. En outre, ces derniers sont beaucoup plus fréquemment conventionnés en secteur 2 (en 2004, ils sont à 51 % en secteur 2 contre 8 % pour les généralistes sans spécialisation).

Ces différences dans les modes d'activité peuvent donc générer des disparités dans le niveau des honoraires. Toutefois, elles ont finalement une faible influence sur la dispersion globale des honoraires des omnipraticiens : si les omnipraticiens conventionnés en secteur 1 ont un volume d'activité bien supérieur à celui des « secteur 2 » (5 060 actes en moyenne par an en 2004 pour les premiers contre 3 312 pour les seconds) ainsi qu'à celui des généralistes ayant une spécialité Mep (2 812 actes par an), leurs honoraires réels diffèrent peu : 119 000 euros en moyenne en 2004 pour les « secteur 1 », contre 123 000 euros pour les « secteur 2 » et 120 000 euros pour les Mep¹⁶.

¹³ Cf. Fivaz C. et Le Laidier, S., 2001. « Une semaine d'activité des généralistes libéraux », Point Stat, n°33, CNAMTS.

¹⁴ Cf. Delattre E. et Dormont B. (2000), pp. 137-161. C'est une estimation de la phase de constitution de la clientèle, évaluée à partir du même panel d'omnipraticiens libéraux, mais uniquement pour les années 1979 à 1993. Actuellement, cette phase est beaucoup plus brève, les médecins atteignant plus rapidement un niveau d'activité stable.

¹⁵ Delattre E. et Dormont B. (2005), pp. 148-150, comparent le taux de croissance de la densité médicale départementale à un « indicateur de pression concurrentielle subie par le médecin ». Sur la période 1988-1993, l'augmentation généralisée de la densité médicale a conduit les médecins à subir un rationnement de leur activité. En revanche, entre 1996 et 2000, même si elle est à un niveau très élevé, la densité médicale diminue chaque année pour plus de la moitié des médecins et l'activité individuelle semble se redresser.

¹⁶ Cf. Audric, S. (2006). « Les généralistes pratiqueraient quasiment le double d'actes de ceux des MEP, mais l'écart disparaît en termes d'honoraires, du fait des nombreux dépassements pratiqués par ces derniers ».

III. Les éléments explicatifs à la baisse des dispersions

1. Facteurs explicatifs étudiés et démarche utilisée

Au cours des vingt dernières années, trois facteurs ont contribué, en particulier, à faire varier la dispersion des honoraires des omnipraticiens (encadré 3) :

- La progression du nombre de femmes médecins et de leur activité
- L'évolution de l'activité des praticiens nouvellement installés.
- L'évolution des densités départementales d'omnipraticiens libéraux

Les femmes et les omnipraticiens en début de carrière appartenant majoritairement au bas de la distribution (qui s'est le plus transformé sur la période), c'est l'évolution de leurs honoraires (et l'homogénéisation intra-groupe et inter-groupe¹⁷) qui explique la majeure partie de la baisse des dispersions observée globalement.

Bien sûr, ces différents facteurs explicatifs se combinent. C'est pourquoi les analyses sur des sous-populations basées sur les indicateurs de dispersion précédemment utilisés (tableaux 2 et 4) seront complétées par des régressions quantiles qui permettent de raisonner toutes choses égales par ailleurs (tableau 3 et annexes 1 et 2).

Le choix des facteurs à prendre en compte dans ces régressions s'inspire des modélisations adoptées pour les fonctions de gains [Mincer (1974)].

Celles-ci utilisent généralement le niveau d'éducation des individus ainsi qu'une fonction quadratique de leur expérience sur le marché du travail pour expliquer le logarithme de leur salaire.

Ici, le travail sur les omnipraticiens rend inutile l'étude des rendements de l'éducation (le nombre d'années d'études théorique étant de neuf années pour tous¹⁸) ; seule l'expérience en libéral¹⁹ peut être utilisée comme mesure du capital humain.

La variable expliquée est le logarithme des honoraires réels.

¹⁷ L'étude sur les dispersions peut être décomposée en dispersions « intra-groupes », c'est à dire au sein d'un même sous-groupe (celui des femmes par exemple) ou en dispersions « inter-groupes », c'est-à-dire entre deux sous-groupes ou plus (celui des hommes et celui des femmes par exemple).

¹⁸ Le nombre d'années d'études des médecins du panel varie en fonction de la période à laquelle le médecin a commencé ses études : il était de sept ans avant la réforme des études médicales de 1984, puis il est passé à neuf ans ensuite. En pratique, ce nombre est généralement compris entre huit et douze ans, du fait des redoublements en PCEM1, au cours de la scolarité ou lors du concours de l'internat. Néanmoins, étant donné que les redoublements ne contribuent pas à accroître le niveau de capital humain du médecin, on supposera que le nombre d'années d'études est identique pour tous et fixé à neuf ans. Il n'est donc pas inclus dans la fonction de gains estimée.

¹⁹ Une autre variable caractéristique de l'ancienneté, l'expérience dans la médecine (calculée à partir de l'année de thèse de l'omnipraticien), aurait également pu être choisie. En plus de l'expérience en libéral, cette deuxième variable prend en compte toute l'expérience de l'omnipraticien avant l'installation à son compte, à savoir l'expérience en tant que médecin remplaçant ou médecin salarié par exemple. On retrouve ici la distinction souvent étudiée dans les fonctions de gains entre l'expérience *transférable* d'une entreprise à l'autre (ici la totalité de l'expérience dans l'exercice de la médecine) et l'expérience *non transférable*, spécifique à une entreprise (ici, à un type d'activité, l'exercice en libéral). C'est cette deuxième variable qui a été retenue.

Étant donné les constats précédents, trois variables explicatives sont introduites : le sexe de l'omnipraticien, une fonction quadratique de son ancienneté et la densité d'omnipraticiens libéraux de son département d'exercice²⁰.

Deux séries de régressions quantiles ont été mises en œuvre.

Les premières sont estimées hommes et femmes confondus, ce qui revient à faire une hypothèse forte : les rendements de l'ancienneté et l'influence de la densité médicale sur le logarithme des honoraires sont identiques pour les deux sexes. Afin de lever cette contrainte, des régressions quantiles par sexe ont donc été estimées : elles indiquent très nettement une influence de l'ancienneté et de la densité médicale qui diffère en fonction du sexe de l'omnipraticien. Les résultats de ces régressions sont présentés en annexe (annexe 1, tableaux 6 et 7).

L'utilisation des régressions quantiles apporte un éclairage supplémentaire par rapport à l'estimation par moindres carrés ordinaires. Alors que ces derniers décrivent comment la moyenne conditionnelle des honoraires dépend des variables explicatives, les régressions quantiles indiquent comment les différents quantiles de la distribution conditionnelle des honoraires dépendent de ces variables. Ainsi, l'influence des facteurs explicatifs peut ne pas être identique à tous les points de la distribution, conduisant à des écarts entre hommes et femmes ou entre omnipraticiens en début de carrière et en régime permanent qui varient le long de la distribution conditionnelle.

Par souci de cohérence avec les analyses descriptives précédentes, seulement cinq régressions quantiles (10^{ème}, 25^{ème}, 50^{ème}, 75^{ème} et 90^{ème}) sont estimées. Le choix de ces cinq régressions quantiles ne permet pas de décrire l'ensemble de la distribution conditionnelle des honoraires : cela aurait nécessité l'utilisation d'une plus large panoplie de régressions quantiles, chacune d'entre elles caractérisant seulement un ou deux points de la distribution conditionnelle.

L'estimation, réalisée pour les cinq régressions simultanément, s'appuie sur le programme de minimisation suivant :

$$\text{Min} \left(\sum_{y_i \geq x_i \beta} \theta |y_i - x_i \beta| + \sum_{y_i < x_i \beta} (1 - \theta) |y_i - x_i \beta| \right),$$

où $\theta = 0,1, 0,25, 0,5, 0,75$ ou $0,9$ selon la régression quantile estimée

Dans un objectif de comparaison des deux méthodes, les résultats obtenus par moindres carrés ordinaires sont également retranscrits.

En outre, quatre années sont retenues pour ces estimations : 1983, 1992, 1998 et 2004.

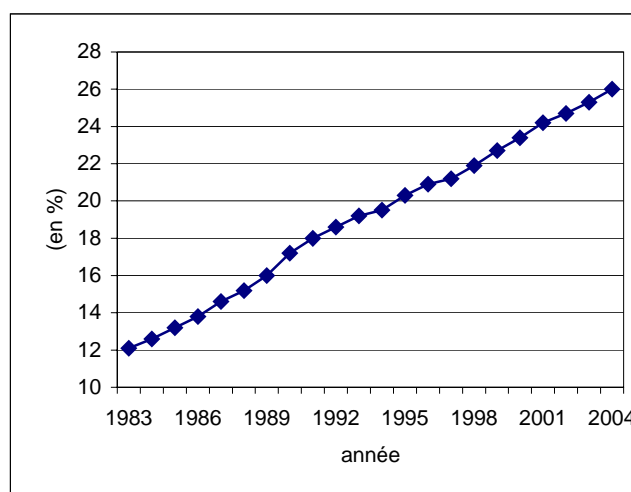
²⁰ La valeur des densités médicales (soit le nombre d'omnipraticiens pour cent mille habitants) départementales pour les années de 1983 à 2004 provient des données du SNIR (CNAMTS).

2. La féminisation des omnipraticiens et la progression de l'activité féminine

A. Un point sur la structure des omnipraticiens par sexe

Passant de 12,1 % en 1983 à 26 % en 2004, la part des femmes au sein des omnipraticiens libéraux a doublé, cette progression étant continue sur la période (graphique 4).

Graphique 4 : Évolution de la part des femmes dans la population des omnipraticiens entre 1983 et 2004



Source : Panel de médecins libéraux du SNIR (CNAMTS), 1983-2004

Il est possible d'expliquer cette féminisation croissante de la profession par les réformes successives des études médicales.

Pour les étudiants des années 1960-1970, le métier de médecin généraliste apparaissait plutôt comme une profession masculine. Les femmes choisissaient en grande majorité des spécialités (par le biais des CES, Certificats d'études spéciales) comme la gynécologie, la pédiatrie, la dermatologie ou étaient plus présentes dans d'autres professions libérales de santé comme les infirmières.

La réforme des études médicales de 1982 a supprimé les CES et scindé les études de médecine entre « résidanat » (pour la médecine générale) et « internat » (pour la médecine spécialisée). Les femmes se sont alors davantage tournées vers la médecine générale, au détriment des spécialités (que la plupart considéraient, entre autres choses, comme trop difficiles). La profession de médecin généraliste s'est donc fortement féminisée²¹.

Aujourd'hui, cette profession reste toujours majoritairement choisie par les hommes, en comparaison avec d'autres professions de santé : en 2004, 91 % des aides-soignants, 87 % des infirmiers, et 64 % des pharmaciens sont des femmes²². Mais la proportion des femmes généralistes est amenée à augmenter très fortement dans le futur : actuellement, les nouveaux étudiants dans les premières années de médecine sont majoritairement des femmes.

²¹ Lire pour plus de détails : HARDY-DUBERNET, A-C.(dir.), ARLIAUD, M., HORELLOU-LAFARGE, C., LE ROY, F et BLANC, M-A. (2001).

²² Ces données sont extraites de BESSIERE S. (2005).

La féminisation accrue de la profession, accompagnée d'une forte progression de l'activité des femmes, a conduit à homogénéiser la profession de médecin généraliste et à diminuer la dispersion des honoraires au sein des omnipraticiennes et entre les deux sexes.

B. Une homogénéisation des honoraires féminins

Entre 1983 et 2004, ce sont les honoraires des omnipraticiennes du bas de la distribution des honoraires féminins qui connaissent la plus forte croissance.

Ce phénomène est lié à l'entrée des femmes dans la profession : le métier de généraliste étant presque exclusivement masculin avant les années 1980, plus de 60 % des femmes présentes en 1983 commençaient à peine leur carrière et étaient, comme tout médecin qui débute, en phase de constitution de clientèle. Au début de la période d'étude, une grande majorité des femmes en exercice possède donc de très faibles honoraires et le bas de la distribution se situe alors à niveau très faible : les honoraires des femmes du premier décile sont de 20 500 euros (43 400 pour les hommes) et ceux des femmes du premier quartile sont de 32 700 euros (66 700 pour les hommes), en euros constants de 2004.

Ce constat permet également d'expliquer la croissance très rapide de leurs honoraires sur la période. Les honoraires des omnipraticiennes du premier décile et du premier quartile des femmes ont augmenté de 115 % et 99 % (en euros constants de 2004) entre 1983 et 2004 (passant de 20 500 à 44 00 euros et de 32 700 à 65 200 euros).

En outre, les autres quantiles de la distribution des honoraires ont tous connu une progression significative : la médiane, le troisième quartile et le neuvième décile ont crû respectivement de 78 %, 62 % et 45 % entre 1983 et 2004, amplifiant le processus d'homogénéisation des honoraires féminins.

La très forte hausse de l'activité²³ féminine entre 1983 et 2004 (+41% en moyenne contre 4 % pour les hommes) peut expliquer cette croissance des honoraires de grande ampleur, qui s'est répercutée sur l'ensemble de la distribution des honoraires féminins.

La restriction des données aux seuls omnipraticiens de plus de sept ans d'ancienneté, afin d'éliminer l'effet « âge » décrit précédemment, renforce ce constat : l'activité des femmes de plus de sept ans d'ancienneté augmente en moyenne de 28 % sur la période, alors qu'elle diminue de 3 % pour les hommes d'ancienneté comparable.

L'amplification du processus de féminisation, seulement amorcé en 1983, ainsi que la progression de l'activité féminine ont donc rendu le groupe des omnipraticiennes beaucoup

²³ Elle est définie ici comme le nombre de rencontres annuelles entre l'omnipraticien et ses patients ; c'est la somme du nombre de consultations, de visites et d'actes cotés en K (actes de chirurgie) par omnipraticien. Cette mesure ne correspond pas exactement à l'activité réelle des omnipraticiens libéraux, dans la mesure où une part importante de leur métier consiste aussi à recevoir des visiteurs médicaux, fournir des conseils par téléphone, assister à des séances de formation et, pour quelques-uns aussi, exercer une activité salariée dans un établissement de santé. La DREES utilise généralement un autre indicateur pour définir l'activité des omnipraticiens, à partir d'une décomposition de la croissance moyenne des honoraires en un effet tarif et un effet volume. L'indice de volume d'activité est donc obtenu en divisant l'indice d'évolution des honoraires (hors dépassements, mais avec les frais de déplacement) par un indice des prix. Ce dernier représente l'évolution des tarifs annuels moyens des actes conventionnés, pondérés par la structure des actes effectués dans l'année étudiée (nombre d'actes pour les consultations et visites, nombre de coefficients multiplicatifs des lettres clés pour les actes techniques). Pour plus de détails, Cf. Legendre, N. (2005).

plus homogène. Alors qu'en 1983, les honoraires des 25 % de femmes les mieux rémunérées étaient 2,2 fois supérieurs à ceux des 25 % de femmes les moins bien rémunérées, ce rapport s'est réduit à 2 fois en 1992 et à 1,8 fois en 2004 (tableau 2).

Chez les femmes, les dispersions observées en haut et en bas de la distribution se sont donc réduites dans des proportions quasiment identiques : passant de 1,46 en 1983 à 1,33 en 2004, le rapport entre le troisième quartile et la médiane a évolué de façon proche du rapport entre la médiane et le premier quartile, passé lui de 1,52 à 1,36.

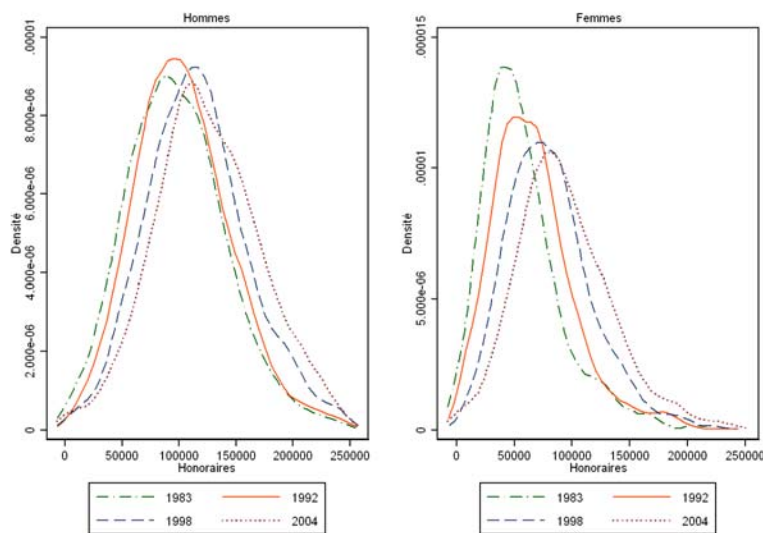
Le déplacement des courbes de densité vers la droite (graphique 5) est d'ailleurs de même ampleur pour le haut et le bas de la distribution²⁴.

Tableau 2 : Évolution de la dispersion des honoraires par sexe entre 1983 et 2004

	Omnipraticiens hommes					Omnipraticiens Femmes				
	83	88	92	98	04	83	88	92	98	04
Q3/Q1	1,88	1,86	1,77	1,69	1,68	2,22	2,16	2,05	1,88	1,80
Q3/Q2	1,32	1,31	1,30	1,26	1,28	1,46	1,42	1,36	1,31	1,33
Q2/Q1	1,43	1,41	1,36	1,35	1,31	1,52	1,52	1,50	1,43	1,36
D9/D1	3,61	3,52	3,16	3,07	2,93	5	5,53	4,27	3,59	3,38
D9/D5	1,64	1,64	1,61	1,58	1,60	2,06	1,94	1,82	1,72	1,68
D5/D1	2,19	2,14	1,97	1,94	1,84	2,43	2,85	2,34	2,09	2,01

Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004

Graphique 5 : Évolution de la distribution des honoraires par sexe entre 1983 et 2004



Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004 (honoraires en euros constants de 2004)

Contrairement à ce qui était observé pour l'ensemble des omnipraticiens, cette homogénéisation des honoraires féminins s'est, en plus, réalisée de manière continue sur la

²⁴ Si le haut de la distribution des honoraires se déplaçait vers la gauche (signe d'une baisse des honoraires des omnipraticiennes les mieux rémunérées), la baisse des dispersions serait très nette graphiquement. Ici, l'ensemble de la distribution des omnipraticiennes se déplace vers la droite (car la progression des honoraires concerne l'ensemble des omnipraticiennes) et la baisse des dispersions pour le haut de la distribution est alors visible par la queue de distribution plus épaisse et moins étalée : une part plus importante d'omnipraticiennes possède des honoraires élevés.

période. Seules les années 1983 à 1988 sont marquées par une alternance de hausse et de baisse des dispersions, les dispersions interdéciles de 5 en 1983 passant même à 5,77 en 1984 et 5,53 en 1988 (du fait de la très forte variabilité des honoraires du bas de la distribution sur cette période).

Chez les hommes médecins, les rapports interquartiles ont également baissé, passant de 1,88 à 1,68. Cette diminution de la dispersion des honoraires en leur sein reste toutefois inférieure à celle constatée pour les femmes.

C. Une réduction des disparités d'honoraires entre les hommes et les femmes

Cette réduction de la dispersion des honoraires au sein des femmes a conduit à une diminution des disparités entre sexes.

En moyenne, toutes choses égales par ailleurs (et en particulier à niveau d'ancienneté identique), les honoraires des femmes étaient de 58 % inférieurs à ceux des hommes en 1983 (tableau 3). Ces écarts moyens ont diminué au cours des années, puisqu'ils sont passés à 47 % en 1992, 37 % en 1998 et 35 % en 2004. Ainsi, même si les honoraires des femmes se maintiennent à un niveau toujours inférieur à celui des hommes (leur activité étant, elle aussi, plus faible²⁵), la très forte progression de leur activité sur la période a favorisé le rapprochement progressif des honoraires des deux sexes²⁶.

Toutefois, les régressions quantiles, qui permettent de se demander pour chaque quantile choisi et sachant les autres variables fixées (et notamment l'ancienneté), dans quelle mesure le logarithme des honoraires des hommes et des femmes diffère, montrent que les résultats précédents ne sont pas identiques à tous les quantiles de la distribution conditionnelle des honoraires²⁷. L'influence du sexe est, en effet, décroissante avec les quantiles considérés (et ce, quelle que soit l'année), la dispersion étant donc plus faible pour les quantiles du haut de la distribution que pour ceux du bas. Ainsi, en 1983, au premier décile de la distribution conditionnelle, les honoraires des femmes étaient de 73 % inférieurs à ceux des hommes, chiffre qui passe à 37 % au neuvième décile.

Par ailleurs, cet écart diminue avec les années : en 2004, il n'est plus que de 42 % au premier décile et de 26 % au neuvième décile. L'influence du sexe de l'omnipraticien perdure donc, à ancienneté donnée : les écarts importants entre hommes et femmes dans les années 1980 ne sont pas qu'un « effet structure » lié à la grande majorité de femmes en début de carrière, et leur évolution sur la période n'est donc pas qu'un effet « début de carrière » qui évolue avec le temps.

²⁵ Elles travaillent moins souvent que les hommes les mercredi, samedi et dimanche, et s'absentent en moyenne un jour de plus au cabinet par semaine [cf. Fivaz, C. et Le Laidier, S. (2000)].

²⁶ Les projections réalisées par X. Niels et J-P. Perret (2002) montrent que si le phénomène de rattrapage se poursuit au même rythme (l'activité des femmes représentait 60 % de celle des hommes en 1989, mais plus de 70 % de celle des hommes en 2000), hommes et femmes auront un niveau d'activité moyen identique vers 2025.

²⁷ Les coefficients obtenus indiquent, en effet, relativement aux hommes, le désavantage des femmes en terme d'honoraires le long de la distribution conditionnelle.

Tableau 3 : Régressions quantiles et par moindres carrés ordinaires du logarithme des honoraires²⁸

	1983						1992					
	MCO	0.1	0.25	0.5	0.75	0.9	MCO	0.1	0.25	0.5	0.75	0.9
Constante (écart-type)	12,206* (0,066)	11,898* (0,194)	11,891* (0,079)	12,151* (0,071)	12,267* (0,056)	12,369* (0,081)	12,007* (0,059)	11,677* (0,144)	11,783* (0,083)	12,032* (0,061)	12,253* (0,051)	12,3* (0,066)
Sexe (écart-type)	-0,580* (0,033)	-0,728* (0,096)	-0,640* (0,042)	-0,572* (0,038)	-0,459* (0,029)	-0,366* (0,048)	-0,466* (0,021)	-0,632* (0,058)	-0,525* (0,033)	-0,45* (0,023)	-0,369* (0,0198)	-0,308* (0,027)
Ancienneté (écart-type)	0,044* (0,004)	0,064* (0,009)	0,066* (0,005)	0,051* (0,005)	0,045* (0,0036)	0,042* (0,005)	0,048* (0,004)	0,075* (0,010)	0,055* (0,004)	0,043* (0,004)	0,034* (0,004)	0,031* (0,004)
Ancienneté ² (écart-type)	-0,0012* (0,0001)	-0,002* (0,0002)	-0,002* (0,0002)	-0,0013* (0,0001)	-0,0011* (0,0001)	-0,001* (0,0001)	-0,001* (0,0001)	-0,002* (0,0003)	-0,0014* (0,0001)	-0,0011* (0,0001)	-0,0009* (0,0001)	-0,0007* (0,0001)
Densité (écart-type)	-0,0064* (0,0005)	-0,009* (0,0016)	-0,006* (0,0005)	-0,0053* (0,0005)	-0,0046* (0,0004)	-0,0043* (0,0006)	-0,0046* (0,0004)	-0,007* (0,001)	-0,0046* (0,0005)	-0,004* (0,0004)	-0,0039* (0,0004)	-0,0028* (0,0005)
R ² (ou Pseudo-R ²)	0,14	0,094	0,118	0,115	0,096	0,066	0,152	0,106	0,108	0,096	0,077	0,054

	1998						2004					
	MCO	0.1	0.25	0.5	0.75	0.9	MCO	0.1	0.25	0.5	0.75	0.9
Constante (écart-type)	12,264* (0,064)	11,855* (0,126)	12,079* (0,091)	12,242* (0,054)	12,365* (0,061)	12,537* (0,078)	12,375* (0,081)	12,047* (0,174)	12,073* (0,085)	12,291* (0,060)	12,408* (0,068)	12,598* (0,067)
Sexe (écart-type)	-0,369* (0,019)	-0,478* (0,036)	-0,428* (0,027)	-0,366* (0,018)	-0,325* (0,018)	-0,277* (0,022)	-0,354* (0,022)	-0,416* (0,044)	-0,362* (0,025)	-0,336* (0,015)	-0,298* (0,018)	-0,264* (0,019)
Ancienneté (écart-type)	0,020* (0,004)	0,032* (0,008)	0,021* (0,005)	0,016* (0,003)	0,015* (0,003)	0,018* (0,0039)	0,018* (0,005)	0,022* (0,009)	0,017* (0,005)	0,013* (0,0038)	0,013* (0,004)	0,014* (0,004)
Ancienneté ² (écart-type)	-0,0005* (0,0001)	-0,0009* (0,0002)	-0,0005* (0,0001)	-0,0004* (0,0001)	-0,00038* (0,0001)	-0,0004* (0,0001)	-0,0005* (0,0001)	-0,0007* (0,0002)	-0,0004* (0,0001)	-0,0004* (0,00009)	-0,0003* (0,0001)	-0,0003* (0,0001)
Densité (écart-type)	-0,005* (0,0005)	-0,006* (0,0009)	-0,005* (0,0006)	-0,004* (0,0004)	-0,0031* (0,0005)	-0,0035* (0,0005)	-0,0046* (0,0005)	-0,006* (0,001)	-0,004* (0,0006)	-0,003* (0,0004)	-0,002* (0,0005)	-0,0027* (0,0005)
R ² (ou Pseudo-R ²)	0,098	0,066	0,074	0,074	0,058	0,053	0,063	0,044	0,056	0,056	0,049	0,043

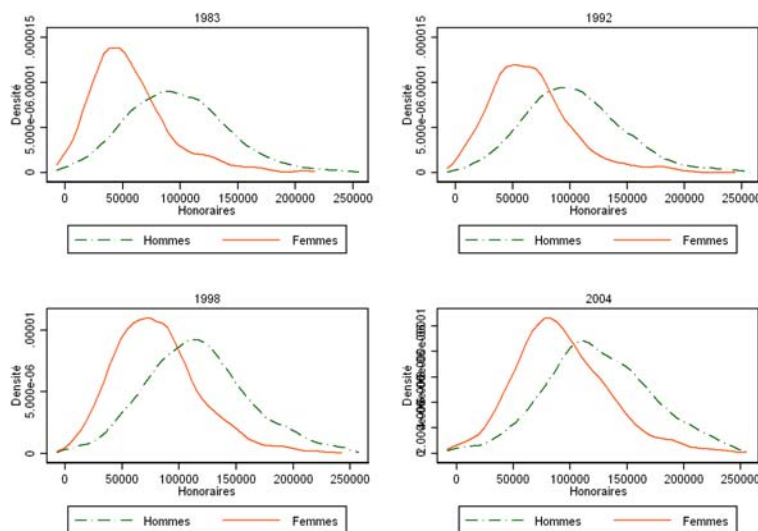
Lecture :

- « * » signifie que les coefficients sont significatifs au seuil de 5 %, et « ** » au seuil de 10 %.
- Colonne « MCO » en 1983 : toutes choses égales par ailleurs, les honoraires des femmes sont en moyenne 58 % plus faibles que ceux des hommes.
- Interpréter les coefficients estimés des différentes régressions quantiles revient à commenter l'influence d'une modification d'une des variables explicatives sur le logarithme des honoraires, au niveau du $\theta^{ème}$ quantile conditionnel, toutes choses égales par ailleurs. Ainsi, pour la 10ème régression quantile en 1983, le coefficient de la variable sexe s'interprète ainsi : au dixième quantile (= premier décile) de la distribution conditionnelle des honoraires, toutes choses égales par ailleurs, les honoraires des femmes sont environ 73 % plus faibles que ceux des hommes.

La distribution des honoraires des omnipraticiens hommes et femmes s'est donc rapprochée au cours des vingt dernières années, même si les femmes continuent à percevoir des honoraires plus faibles (graphique 6).

²⁸ Les écarts types sont estimés en utilisant 1000 itérations Bootstrap. Les estimateurs obtenus sont convergents et asymptotiquement normaux, ce qui autorise la réalisation de tests de significativité de ces estimateurs ainsi que de tests de différence entre les estimateurs obtenus aux différents quantiles (c'est-à-dire à différents points de la distribution des honoraires). Toutes ces procédures sont décrites dans l'annexe 2.

Graphique 6 : Évolution, sur la période 1983-2004, des écarts entre la distribution des honoraires hommes et des femmes



Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004 (honoraires en euros constants de 2004)

Cette première mise en œuvre des régressions quantiles permet de comprendre tout leur intérêt (annexe 1).

Les régressions interdéciles et interquartiles (annexe 1, tableau 11) estiment un coefficient de la variable « sexe » significativement différent de zéro pour l'ensemble des années, signe que l'influence de cette caractéristique est significativement différente entre les quantiles. De plus, les graphiques des coefficients estimés (annexe 1, graphiques 12 et 13), indiquent que l'influence du sexe n'est pas constante à tous les points de la distribution conditionnelle du logarithme des honoraires. Enfin, ces mêmes graphiques montrent également que les coefficients estimés par les régressions quantiles sortent, la plupart du temps, de l'intervalle de confiance défini par les moindres carrés ordinaires.

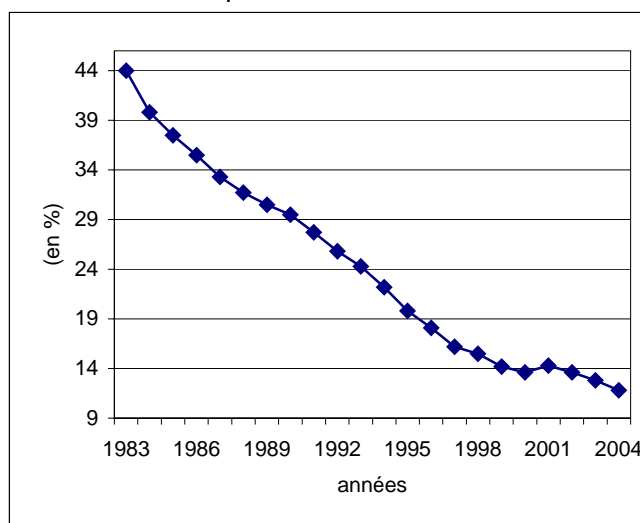
Ces trois éléments justifient l'utilisation de la méthode des régressions quantiles : elle apporte une information supplémentaire par rapport à celle transmise par les moindres carrés ordinaires. En effet, l'influence de la variable « sexe » diffère en fonction du quantile estimé et la valeur trouvée « en moyenne » n'est pas une mesure suffisante de l'influence de cette variable sur l'ensemble de la distribution conditionnelle du logarithme des honoraires.

3. La progression de l'activité des omnipraticiens en début de carrière

A. Un point sur la structure des omnipraticiens par âge

Passant de 45 % en 1983 à 11,8 % en 2004, la part des omnipraticiens de moins de sept années d'ancienneté (et considérés ici comme en début de carrière) a fortement diminué sur la période (graphique 7).

Graphique 7 : Évolution de la part des omnipraticiens de moins de sept années d'ancienneté dans la population des omnipraticiens, entre 1983 et 2004



Source : Panel de médecins libéraux du SNIR (CNAMTS), 1983-2004

Ce vieillissement est la conséquence des politiques gouvernementales en matière de Numerus Clausus. Celui-ci, instauré en 1971 pour limiter le nombre d'étudiants admis à s'inscrire en deuxième année de médecine, est à l'origine d'une diminution de nouveaux installés chaque année. Fixé à près de 8 600 à ses débuts, il a été réduit régulièrement jusqu'en 1993 (à un niveau de 3 500), avant d'être de nouveau rehaussé, à partir de 1999 et de manière significative en 2002. En 2004, il était de 5 500.

Il n'est donc pas encore possible de voir les conséquences de cette hausse sur le rajeunissement de la population des omnipraticiens (il faudra attendre 2010 environ pour constater ses effets sur la structure par âge des omnipraticiens).

Comme la féminisation, ce vieillissement de la population des omnipraticiens est à l'origine d'une diminution de la dispersion globale des honoraires.

B. Une homogénéisation des honoraires des omnipraticiens en début de carrière

Les honoraires réels des omnipraticiens en début de carrière ont en effet très fortement augmenté sur la période et en leur sein, ce sont les honoraires les plus faibles qui ont connu la progression la plus spectaculaire (graphique 8), contribuant à homogénéiser la situation des omnipraticiens nouvellement installés : la valeur de leur premier décile (quartile) d'honoraires est passée de 32 500 (51 000) euros en 1983 à 50 700 (72 700) euros en 2004, soit une hausse de 56 % (43 %). Dans le même temps, les valeurs de leur troisième quartile et de leur neuvième décile ont toutes deux augmenté de 22 % environ. La progression de l'activité (et des honoraires) a en effet plus largement profité à ceux pour lesquels elle était auparavant très faible.

Le ralentissement de la croissance de la densité médicale, suite à la diminution du nombre de nouveaux omnipraticiens, et la hausse de la demande de soins ont induit une moindre concurrence entre les omnipraticiens nouvellement installés.

Les évolutions de carrière des jeunes praticiens s'en sont trouvées modifiées²⁹. En effet, leur phase d'installation est devenue plus courte puisqu'ils ont pu bénéficier, dès leurs premières années, d'une clientèle plus nombreuse. *A contrario*, l'afflux de nombreux jeunes omnipraticiens dans les années 1970 et 1980 avait été à l'origine d'une stagnation des honoraires des promotions installées jusque dans les années 1990.

Dès lors, un omnipraticien en début de carrière dans les années 1990 et 2000 avait peu en commun avec son homologue des années 1980 : en 2004, ses honoraires réels étaient, en moyenne, de 29 % plus élevé qu'en 1983 (en euros constants de 2004).

De ce fait, la dispersion des honoraires des omnipraticiens en début de carrière s'est fortement réduite au cours des années (tableau 4). Alors que les honoraires des 10 % les mieux rémunérés étaient, en 1983, 4,3 fois supérieurs aux honoraires des 10 % les moins bien rémunérés, ce rapport est passé à 4 en 1992 et à 3,4 en 2004 (principalement grâce à la contraction de la dispersion du bas de la distribution).

Tableau 4 : Évolution de la dispersion des honoraires par ancienneté entre 1983 et 2004

	Omnipraticiens de moins de sept ans d'ancienneté					Omnipraticiens de plus de sept ans d'ancienneté				
	83	88	92	98	04	83	88	92	98	04
Q3/Q1	2,13	2,14	2,07	1,94	1,84	1,87	1,86	1,83	1,79	1,79
Q3/Q2	1,39	1,43	1,37	1,34	1,32	1,30	1,31	1,31	1,29	1,31
Q2/Q1	1,53	1,49	1,51	1,45	1,40	1,43	1,41	1,39	1,38	1,36
D9/D1	4,3	4,57	4,07	3,56	3,39	3,72	3,58	3,43	3,34	3,27
D9/D5	1,79	1,86	1,76	1,65	1,68	1,62	1,65	1,64	1,63	1,66
D5/D1	2,41	2,45	2,32	2,15	2,02	2,29	2,17	2,08	2,04	1,97

Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004

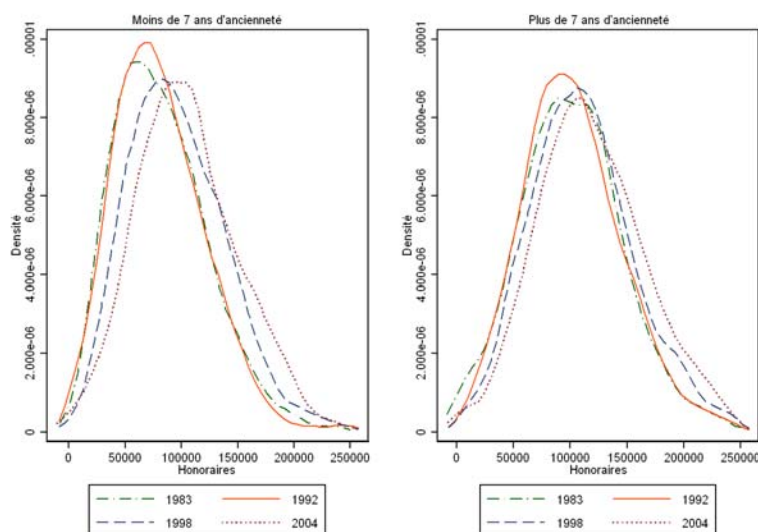
Le déplacement vers la droite de la distribution des honoraires entre 1983 et 2004 (graphique 8) a donc été très marqué pour les omnipraticiens en début de carrière, notamment à partir de 1992, date à laquelle le ralentissement de la progression de la densité médicale commence à faire sentir ses effets. Ce constat se retrouve aussi à l'examen des écarts interquartiles puisque c'est à partir de 1992 que les dispersions diminuent à un rythme régulier (les rapports interquartiles passent par exemple de 2,07 en 1992 à 1,94 en 1998 et 1,84 en 2004).

Cette réduction de la dispersion des honoraires est beaucoup moins marquée pour les omnipraticiens exerçant en « régime permanent »³⁰, assimilés ici aux omnipraticiens de plus de sept ans d'ancienneté. La progression de leurs honoraires a en outre été sensiblement plus faible : les honoraires des praticiens du premier décile ont augmenté de 35 %, passant de 43 600 euros en 1983 (en euros constants de 2004) à 59 000 euros en 2004, ceux des médecins du premier quartile de 22 % et de 17 % pour les autres quantiles.

²⁹ L'étude de V. Lucas-Gabrielli et M.J. Sourty-Le Guellec (2004) réalisée à partir des mêmes données sur l'évolution de la carrière des médecins généralistes du secteur I, montre que l'activité des médecins installés dans les années 1990 est, dès le début, plus élevée que celle des médecins ayant commencé à exercer dans les années 1980. De plus, elle se maintient à un niveau supérieur, quel que soit le nombre d'années d'ancienneté.

³⁰ L'expression d'omnipraticien en « régime permanent » est tirée de Delattre E. et Dormont B., 2000. « Induction de la demande de soins par les médecins libéraux français », *Économie et Prévision*, n° 142, pp. 137-161. Elle est définie par opposition aux omnipraticiens « en début de carrière » dont l'ancienneté est inférieure ou égale à sept ans.

Graphique 8 : Évolution de la distribution des honoraires par ancienneté entre 1983 et 2004



Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004 (honoraires en euros constants de 2004)

C. Une réduction des disparités d'honoraires entre les omnipraticiens de moins et de plus de sept ans d'ancienneté

D'après les régressions quantiles (tableau 3), qui mesurent l'influence de l'ancienneté (en prenant en compte son effet quadratique³¹) sur les différents quantiles de la distribution conditionnelle des honoraires, la distribution des omnipraticiens en début de carrière s'est rapprochée de celle des omnipraticiens « en régime permanent » entre 1983 et 2004.

En 1983, l'influence de l'ancienneté diminuait avec les quantiles et variait fortement en fonction de l'ancienneté déjà acquise par l'omnipraticien. Une année d'exercice supplémentaire apportait environ 5,8 % d'honoraires en plus aux omnipraticiens de deux ans d'ancienneté du premier quartile de la distribution conditionnelle, contre 2,6 % pour ceux de dix ans d'ancienneté et 0,6 % pour ceux de quinze ans d'ancienneté. Au troisième quartile de la distribution conditionnelle, ces chiffres étaient respectivement de 4,1 %, 3 % et 1,4 % (annexe 1, tableau 8).

Ces résultats montrent que l'ancienneté a une influence croissante, mais à taux décroissant, sur le niveau des honoraires : une année d'exercice supplémentaire a, en effet, plus d'impact sur les honoraires d'un omnipraticien en début de carrière que sur ceux d'un omnipraticien déjà installé, avec une clientèle régulière.

Jusqu'à la fin des années 1990, l'inégale influence de l'ancienneté sur l'ensemble de la distribution des honoraires a donc été à l'origine d'une partie de la dispersion entre les quantiles : en 1992, une année d'exercice supplémentaire accroissait les honoraires des

³¹ Si le $\theta^{\text{ème}}$ quantile de la distribution conditionnelle des honoraires est ainsi défini : $Quant_{\theta}(y|x) = \alpha(\theta) + \beta_1(\theta) \text{sexe} + \beta_2(\theta) \text{anc} + \beta_3(\theta) \text{anc}^2 + \beta_4(\theta) \text{dens}$, sa dérivée par rapport à l'ancienneté, qui permet de mesurer le rendement de l'ancienneté, est alors $r = \beta_2(\theta) + 2\beta_3(\theta) \text{anc}$. Il dépend donc de la valeur de l'ancienneté à laquelle on l'évalue.

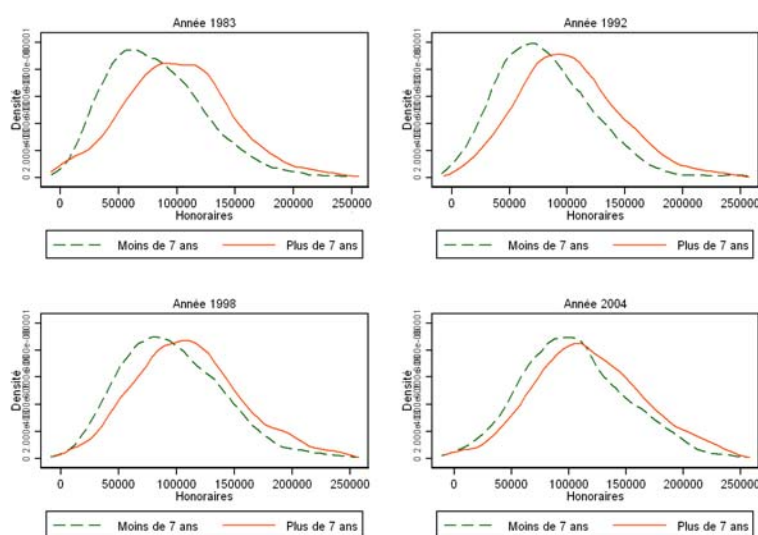
omnipraticiens (de douze ans d'ancienneté) du premier quartile de la distribution conditionnelle de 2,1 %, contre 1,7 % pour ceux des omnipraticiens à la médiane et 1,2 % pour ceux appartenant au troisième quartile.

Toutefois, à partir de 1998, l'influence de l'ancienneté est beaucoup plus faible, identique à tous les quantiles, et n'est donc plus à l'origine de dispersions d'honoraires. D'ailleurs, les coefficients estimés de l'ancienneté et de l'ancienneté au carré deviennent non significativement différents entre les quantiles et ne sortent plus de l'intervalle de confiance défini par les moindres carrés ordinaires (annexe 1, graphiques 16 à 19 et tableau 8). Le coefficient estimé par cette dernière résume donc de façon satisfaisante l'influence d'une année d'ancienneté supplémentaire sur l'ensemble de la distribution conditionnelle du logarithme des honoraires.

Par exemple, l'influence d'une année d'exercice supplémentaire sur la distribution conditionnelle des honoraires varie entre 1 % et 0,8 % (pour des omnipraticiens de douze années d'ancienneté) selon qu'on considère le premier décile ou le neuvième décile de cette distribution. Il n'y a donc plus d'écart significatif dû à l'ancienneté entre les distributions d'honoraires des omnipraticiens, et notamment entre celles des omnipraticiens en début de carrière et en régime permanent.

Ces résultats s'observent graphiquement (graphique 9) : les distributions se resserrent continuellement entre 1983 et 2004. Mais c'est réellement à partir de l'année 1998, et surtout en 2004, que les omnipraticiens en début de carrière et en régime permanent connaissent des distributions très proches, à la fois en terme de niveau et de dispersion.

Graphique 9 : Évolution, sur la période 1983-2004, des écarts entre la distribution des honoraires des omnipraticiens de moins et de plus de sept ans d'ancienneté



Source : Panel d'omnipraticiens libéraux du SNIR (CNAMTS), 1983-2004 (honoraires en euros constants de 2004)

Plus dans le détail, lorsque les régressions quantiles sont effectuées par sexe, l'influence de l'ancienneté diffère entre hommes et femmes (annexe 1, tableaux 6 et 7 pour les coefficients estimés et tableaux 9 et 10 pour le calcul des rendements de l'ancienneté).

Alors que pour les hommes, l'influence de l'ancienneté décroît avec les quantiles, cette dernière est en forme de « U » chez les femmes : elle est élevée pour le bas de la distribution conditionnelle des honoraires (de 1,6 % pour les femmes de douze années d'ancienneté du premier quartile en 1983), puis elle diminue (à 1,5 % pour ces mêmes femmes à la médiane) et croît de nouveau avec le haut de la distribution (influence de 2 % pour le troisième quartile). Toutefois, il apparaît que cette influence est statistiquement la même à tous les quantiles de la distribution des honoraires féminins. Ce même constat n'est vrai pour les hommes qu'à partir de 1998, conformément à la tendance de l'échantillon observée hommes et femmes confondus.

4. L'évolution de la densité médicale dans le département ou la région d'exercice de l'omnipraticien

A. Un point sur l'évolution de la densité médicale entre 1983-2004

Évolution de la densité médicale entre 1983-2004

La densité moyenne d'omnipraticiens libéraux a cru très rapidement entre 1983 et 2004 : de 90 omnipraticiens pour 100 000 habitants, elle est passée à près de 107 en 1997, puis a faiblement diminué les années suivantes pour passer à 103 en 2004³².

Deux périodes peuvent être distinguées (graphique 10) :

1983-1997 se caractérise par une augmentation ininterrompue de la densité médicale.

1997-2004 par une diminution, mais de faible ampleur, de la densité moyenne d'omnipraticiens libéraux, cette dernière se maintenant toutefois à un niveau plus élevé qu'au début des années 1990.

Cette diminution de la densité médicale ne signifie pas que le nombre d'omnipraticiens en activité a diminué (graphique 10). En effet, ce dernier a (presque) continuellement augmenté sur la période : si près de 44 000 omnipraticiens exerçaient en libéral en 1983, ils étaient plus de 58 000 en 2004. Toutefois, la croissance des effectifs s'est stabilisée au cours des années 1990. C'est le moment à partir duquel les conséquences de la diminution du *numerus clausus* se ressentent réellement, puisque le nombre de nouveaux omnipraticiens est de plus en plus faible chaque année.

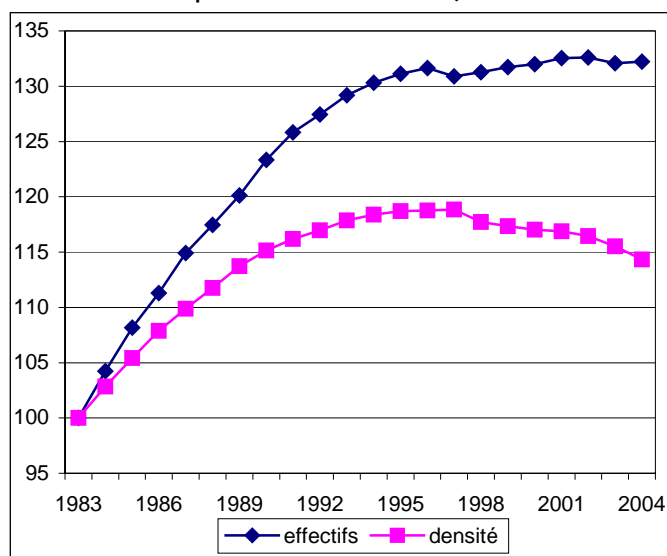
Pour expliquer la diminution de la densité moyenne d'omnipraticiens à partir de 1997, il est possible d'invoquer les effets du Mécanisme d'incitation à la cessation d'activité (MICA), système de préretraite spécifique aux médecins libéraux, créé en 1988.

À ses débuts, il était réservé aux médecins de plus de soixante ans ; ces derniers recevaient une Allocation de remplacement (ADR) jusqu'à soixante-cinq ans, âge auquel ils pouvaient prétendre à leur retraite à taux plein.

³² À titre de comparaison, la densité médicale, toutes spécialités confondues, est de 335 en 2002.

En 1996, ce système a été étendu à tous les médecins de plus de cinquante-sept ans, ce qui peut expliquer un nombre de sorties plus important les années suivantes, conduisant à un ralentissement, voire une diminution de la densité médicale. Le MICA a été fermé aux médecins libéraux en octobre 2003.

Graphique 10 : Évolution des effectifs d'omnipraticiens libéraux et de la densité médicale moyenne (nombre d'omnipraticiens libéraux pour 100 000 habitants) entre 1983 et 2004, base 100 en 1983



Source : Données du SNIR (CNAMTS), 1983-2004

La faible diminution de la densité médicale moyenne cache en réalité de fortes disparités dans la valeur des densités régionales.

L'inégale répartition des omnipraticiens sur le territoire national

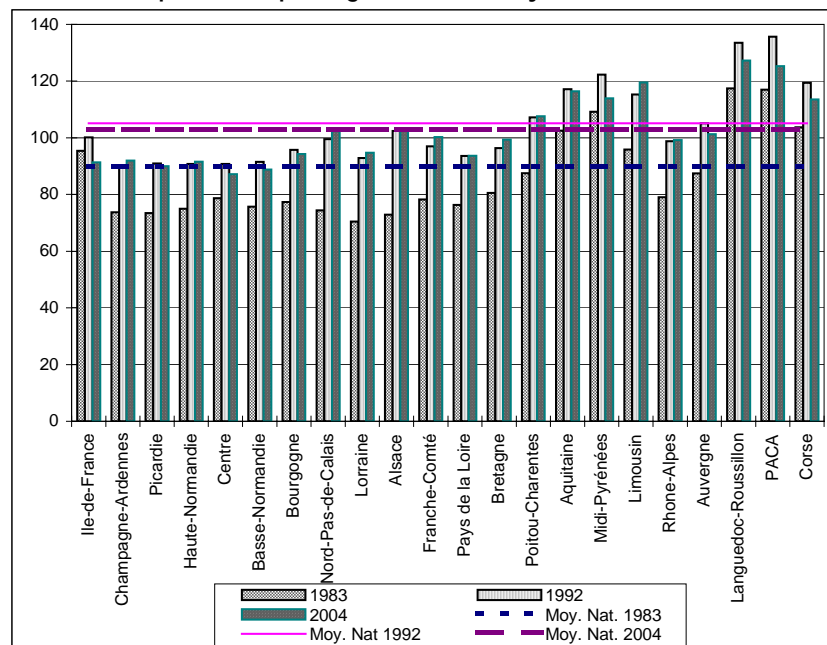
En fonction des régions, les densités d'omnipraticiens peuvent être très contrastées. Ainsi, certaines régions (graphique 11) sont plus fortement dotées en omnipraticiens que d'autres.

Ce phénomène est une des conséquences de la liberté d'installation des omnipraticiens (du moins jusqu'à une date récente) une fois leurs études terminées. Les zones rurales et périurbaines dites « difficiles » se trouvent alors souvent désertées, tout comme certaines régions (Picardie, Centre, Normandie, Champagne-Ardennes, Lorraine, Alsace). D'autres, en revanche, sont très prisées (Languedoc-Roussillon et PACA arrivent très largement en tête). Il y a ainsi deux fois plus d'omnipraticiens dans les régions méridionales qu'au nord de la Loire.

Une deuxième explication tient aux modalités du concours de médecine. En effet, jusqu'à la réforme de 2004 et la mise en place des Épreuves classantes nationales (ECN) pour lutter contre cette inégale répartition des omnipraticiens sur le territoire, le concours de l'internat ne prévoyait pas un nombre de places réservé à la médecine générale. Ce quota dépendait, pour chaque région, du succès des étudiants au concours de l'internat. Les régions dans lesquelles la réussite au concours était très élevée se voyaient plus faiblement dotées en médecins généralistes que d'autres régions, dans lesquelles le succès au concours était plus faible.

La nouvelle réforme a modifié cette répartition, puisqu'un nombre de places fixe pour chaque université de chaque région est attribué à la médecine générale, devenue une spécialité à part entière.

Graphique 11 : L'inégale répartition des omnipraticiens sur le territoire national : évolution de la valeur des densités d'omnipraticiens par région et de la moyenne nationale entre 1983 et 2004



Source : Données du SNIR (CNAMTS), 1983-2004

Toutefois, la répartition des omnipraticiens sur le territoire national s'est modifiée sur une partie de la période considérée.

Entre 1983 et 1992, la densité médicale augmente dans toutes les régions, sans exception. Conformément à l'observation « en moyenne » (graphique 10), toutes les régions connaissent une hausse de leur densité d'omnipraticiens en exercice (graphique 11).

La faible diminution « en moyenne » de la densité médicale à partir de 1997 cache cependant des disparités.

Les régions les plus densément peuplées en omnipraticiens dans les années 1990 voient effectivement leurs densités médicales diminuer. C'est le cas, par exemple, en Languedoc Roussillon (de 133 à 127), Provence-Alpes-Côte d'Azur (de 135 à 125) ou Midi-Pyrénées (de 122 à 114).

En revanche, ces dernières ont plutôt tendance à augmenter dans des régions moins « médicalisées », comme dans le Nord-Pas de Calais (de 99 à 102), en Alsace (de 102 à 103) ou en Franche-Comté (de 97 à 100).

La combinaison de ces deux dynamiques contraires a contribué à diminuer les écarts de densité médicale entre les régions.

Notons que ces évolutions ne modifient toutefois pas la hiérarchie entre les régions, comme si cette répartition était structurelle : les plus densément peuplées en omnipraticiens dans les années 1980 le sont toujours en 2004, et inversement.

B. Les conséquences en terme de dispersion des honoraires

Cette inégale répartition des omnipraticiens sur le territoire national a des conséquences sur la dispersion des honoraires.

Le niveau de la densité médicale locale est source de dispersion entre les omnipraticiens (encadré 3) dans la mesure où elle influence fortement la demande potentielle qui s'adresse à l'omnipraticien. Plus le nombre d'omnipraticiens installés sur une aire géographique est important, plus l'activité et les honoraires de chaque omnipraticien sont potentiellement contraints par une faible demande / une offre abondante.

À titre d'exemple, considérons six régions type³³, classées en trois groupes en fonction de la valeur de leur densité en omnipraticiens³⁴.

GROUPE 1 : deux régions représentatives des régions à forte densité médicale (quelle que soit l'année, la valeur de la densité médicale est très supérieure à la moyenne nationale) : **Languedoc-Roussillon et PACA.**

GROUPE 2 : deux régions représentatives des régions à densité médicale « moyenne », c'est-à-dire dans lesquelles la densité médicale est proche de la moyenne nationale, quelle que soit l'année : **Poitou-Charentes et Auvergne.**

GROUPE 3 : deux régions représentatives des régions à faible densité médicale (la valeur de la densité médicale est très inférieure à la moyenne nationale, pour toutes les années) : **Lorraine et Picardie.**

Tableau 5 : Évolution de la dispersion des honoraires entre 1983 et 2004
par groupe de densité médicale régionale

	GROUPE 1				GROUPE 2				GROUPE 3			
	1983	1992	1998	2004	1983	1992	1998	2004	1983	1992	1998	2004
Densité médicale (1)	117,2	134,6	130,9	126,3	87,5	106,1	107,7	104,3	71,9	91,9	93,9	92,3
Moyenne Nationale	89,9	105,1	105,8	102,8	89,9	105,1	105,8	102,8	89,9	105,1	105,8	102,8
Rapports interquartiles												
Q3/Q1	2,24	1,98	1,98	1,96	1,88	1,77	1,74	1,85	1,80	1,90	1,77	1,65
Q3/Q2	1,41	1,46	1,51	1,44	1,30	1,29	1,28	1,31	1,34	1,34	1,29	1,26
Q2/Q1	1,59	1,36	1,31	1,36	1,45	1,37	1,35	1,41	1,35	1,42	1,38	1,30
Rapports interdéciles												
D9/D1	4,86	4,25	3,95	4,07	3,55	2,98	3,03	3	3,49	3,03	3,06	2,85
D9/D5	1,80	1,80	1,67	1,73	1,50	1,54	1,59	1,58	1,58	1,57	1,61	1,59
D5/D1	2,69	2,36	2,37	2,36	2,36	1,93	1,90	1,89	2,21	1,93	1,89	1,79

Note de lecture :

(1) : il s'agit d'une densité médicale moyenne pour le groupe, c'est-à-dire pour les deux régions sans distinction.

³³ Seules six régions ont été conservées pour faciliter les interprétations. Mais il est possible de généraliser l'analyse à l'ensemble des régions.

³⁴ La hiérarchie des régions en fonction de la valeur de la densité médicale est globalement stable au cours du temps. Ainsi, une région appartenant au premier groupe en 1983 appartiendra généralement à ce même groupe en 2004.

Le niveau de la demande potentielle influence fortement les honoraires ainsi que leur dispersion au sein d'un groupe. Ainsi, la dispersion des honoraires au sein du groupe 1 est toujours plus forte que celle du groupe 2, qui est elle-même supérieure à celle du groupe 3. À titre d'exemple, en 2004, les rapports inter-quartiles sont de 1,96 pour le groupe 1, contre 1,85 pour le groupe 2 et 1,65 pour le groupe 3 (tableau 5).

Plus dans le détail, il est possible d'expliquer l'ampleur de la dispersion au sein de chaque groupe.

Le groupe 1, composé de régions dans lesquelles la densité médicale est plus forte que la moyenne, se caractérise par une forte « concurrence » entre les omnipraticiens car la demande potentielle qui s'adresse à chaque praticien est plus faible que dans d'autres régions.

Contrairement aux deux autres groupes, il existe de très forts contrastes entre les omnipraticiens de ce groupe. En effet, deux hypothèses sont possibles concernant les praticiens de ces régions :

Soit ils ont constitué leur clientèle de manière durable en parvenant à la fidéliser, de telle sorte qu'ils sont assez peu sensibles à la « concurrence »

Soit ils connaissent des difficultés à maintenir une clientèle régulière.

En conséquence, la dispersion des honoraires compris entre le bas et le milieu de la distribution est très forte, puisque les honoraires de certains omnipraticiens (qui connaissent une faible demande) peuvent descendre à un niveau très bas. En 1983 par exemple, le rapport entre le premier décile et la médiane est de 2,69 dans ce groupe contre 2,21 dans le groupe 3.

La dispersion des honoraires compris entre le milieu et le haut de la distribution est également très élevée.

En revanche, les omnipraticiens des groupes 2 et 3 font face à une concurrence plus faible. De ce fait, même si la dispersion des honoraires au sein du groupe 2 est globalement plus élevée que celle du groupe 3, l'écart entre les deux groupes n'est pas aussi significatif qu'avec le groupe 1.

Ce sont toutefois les omnipraticiens du groupe 3 qui connaissent la plus faible dispersion des honoraires du bas de la distribution : le rapport entre les honoraires des 10 % les moins bien rémunérés et la médiane est de 2,21 en 1983 et 1,79 en 2004, contre respectivement 2,36 et 1,89 pour le groupe 2 et 2,69 et 2,36 pour le groupe 1. Ainsi, le plus faible nombre d'omnipraticiens (relativement à la population de la région) et la nécessité de répondre à une demande relativement plus forte que dans les autres groupes, garantirait (même si répondre à la demande peut également représenter une certaine contrainte) aux omnipraticiens de la région d'avoir une activité et donc des honoraires « minimums », ce qui n'est pas toujours le cas pour les omnipraticiens des autres groupes.

En revanche, la dispersion des honoraires du haut de la distribution est proche pour les groupes 2 et 3, beaucoup plus faible que pour le groupe 1 : entre 1983 et 2004, le rapport entre la médiane et le 3ème quartile se maintient aux alentours de 1,30 pour ces deux groupes alors qu'il est supérieur à 1,4 pour les omnipraticiens du groupe 1.

Globalement, et quel que soit le groupe considéré, la dispersion des honoraires s'est réduite entre 1983 et 2004. Ainsi, pour le groupe 1, le rapport interquartile est passé de 2,24 à 1,96 sur la période, ce même rapport déclinant de 1,88 à 1,85 pour le groupe 2 et de 1,80 à 1,65 pour le groupe 3. Cette constatation peut s'expliquer par les facteurs évoqués jusqu'à présent, à savoir le rapprochement des honoraires entre les sexes et entre niveaux d'ancienneté différents, ainsi que la hausse généralisée de la demande de soins. Ces phénomènes, qui touchent l'ensemble des omnipraticiens, ne sont pas spécifiques à une région particulière et peuvent donc expliquer la baisse des dispersions au sein de chaque groupe considéré.

C. Une réduction des disparités d'honoraires entre praticiens exerçant dans des régions inégalement peuplées en omnipraticiens

En fonction de leur région ou département d'exercice, les omnipraticiens font face à une concurrence plus ou moins élevée qui influence fortement le niveau de leurs honoraires. L'inégale répartition des omnipraticiens sur le territoire national est donc source de dispersions d'honoraires entre les omnipraticiens, ce que les régressions quantiles confirment.

Les coefficients étant négatifs et décroissants (en valeur absolue) avec les quantiles conditionnels de la distribution des honoraires, il apparaît qu'une hausse du nombre d'omnipraticiens en activité dans un département diminue plus fortement les honoraires des omnipraticiens les moins bien rémunérés³⁵. En effet, les faibles honoraires de ces derniers sont peut-être la conséquence d'un niveau d'activité déjà contraint par une offre forte.

Ainsi, en 1983, une progression de la densité d'omnipraticiens en activité diminue de 9,7 % les honoraires des praticiens au premier décile de la distribution conditionnelle, contre - 5,7 % pour ceux à la médiane conditionnelle et - 4,9 % pour ceux au neuvième décile (tableau 3).

L'influence de la densité médicale diminue avec les années et elle a moins d'impact sur les dispersions d'honoraires en 1998 et 2004. En 2004 par exemple, elle reste élevée pour les omnipraticiens au premier décile de la distribution conditionnelle (de -6,1 %) mais elle n'est que de - 2,9 % pour les praticiens à la médiane conditionnelle et -2,6 % pour ceux situés au neuvième décile conditionnel. Cela est à mettre en relation avec sa plus faible progression au cours des dernières années de la période d'étude (voire sa diminution dans certains départements), alors que les années 1980 et le début des années 1990 avaient été marquées par son augmentation dans quasiment tous les départements d'exercice.

Toutefois (annexe 1, graphiques 20 et 21) les coefficients estimés par les régressions quantiles pour les années 1983 à 2004 restent globalement en dehors de l'intervalle de confiance défini par les moindres carrés ordinaires et les coefficients sont, dans l'ensemble, significativement différents entre les quantiles (annexe 1, tableau 11). Ces éléments justifient encore l'utilisation des régressions quantiles pour l'étude des dispersions d'honoraires.

³⁵ Les travaux de E. Delattre et B. Dormont sur le comportement des médecins libéraux et les phénomènes de demande induite montrent que la réaction des omnipraticiens à une baisse de la densité médicale n'est pas symétrique à celle observée en cas de hausse : leur activité augmente beaucoup plus lorsque la densité médicale baisse, qu'elle ne diminue dans le cas contraire. Cet effet ne peut pas être testé avec la spécification utilisée ici.

IV. Quelques pistes à explorer

Seules trois variables explicatives significatives ont été conservées ; ce sont par ailleurs celles qui influencent le plus les dispersions d'honoraires. Mais d'autres variables pourraient être incluses dans les régressions :

Le secteur de conventionnement.

Le type d'activité (exercice en libéral à temps plein ou à temps partiel).

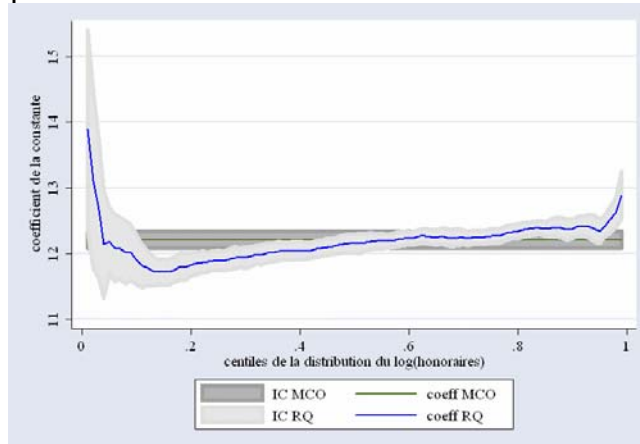
Le mode d'exercice (présence ou non d'un mode d'exercice particulier).

Des éléments caractérisant la population du département d'exercice de l'omnipraticien, notamment son revenu moyen.

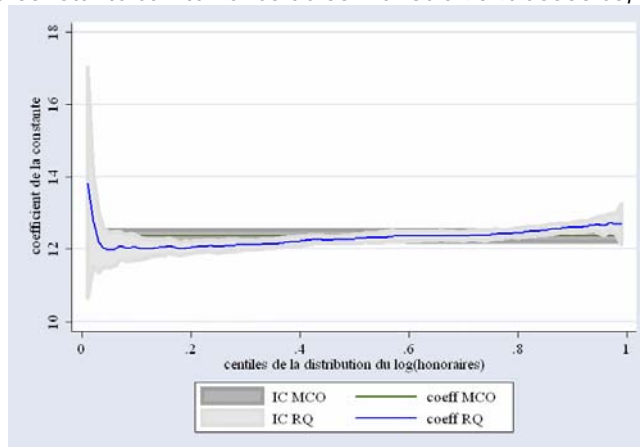
V. Annexe 1 : Résultats complémentaires sur les régressions quantiles

1. Graphiques des coefficients estimés et intervalles de confiance associés

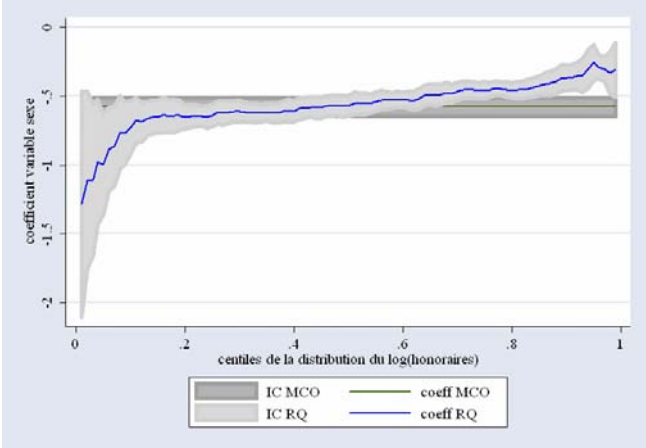
Graphique 12 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la constante et intervalles de confiance à 95 % associés, année 1983



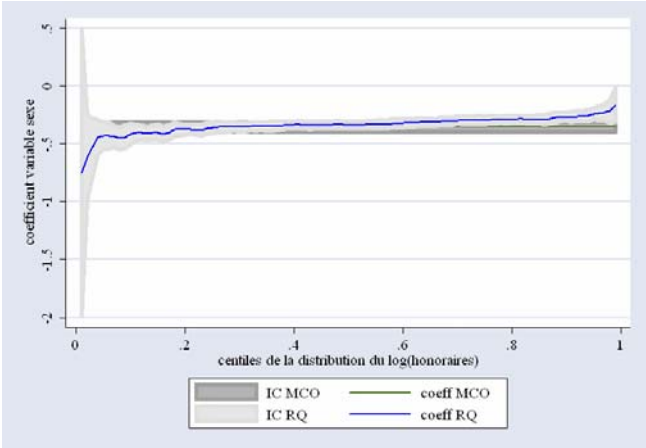
Graphique 13 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la constante et intervalles de confiance à 95 % associés, année 2004



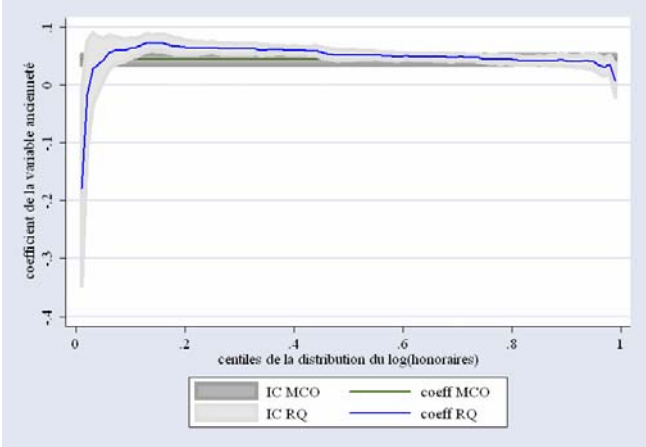
Graphique 14 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « sexe » et intervalles de confiance à 95 % associés, année 1983



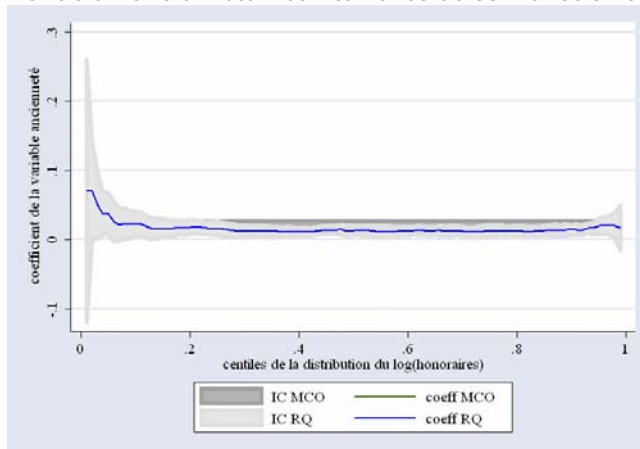
Graphique 15 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « sexe » et intervalles de confiance à 95 % associés, année 2004



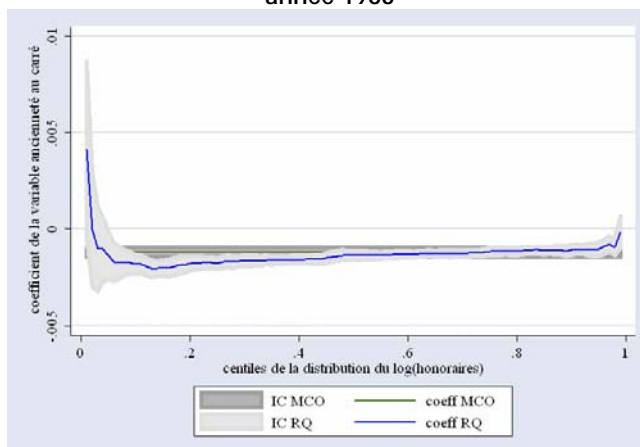
Graphique 16 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « ancienneté » et intervalles de confiance à 95 % associés, année 1983



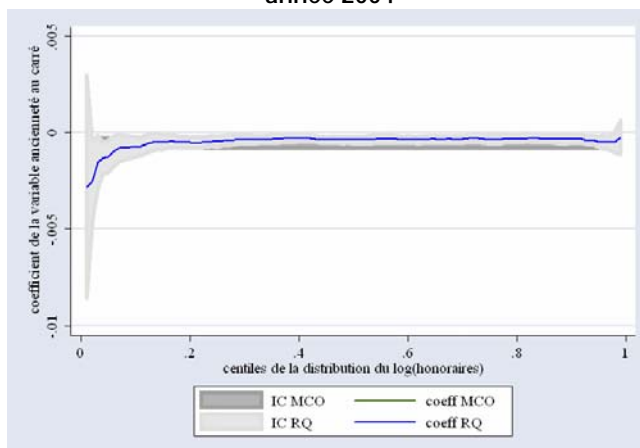
Graphique 17 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « ancienneté » et intervalles de confiance à 95 % associés, année 2004



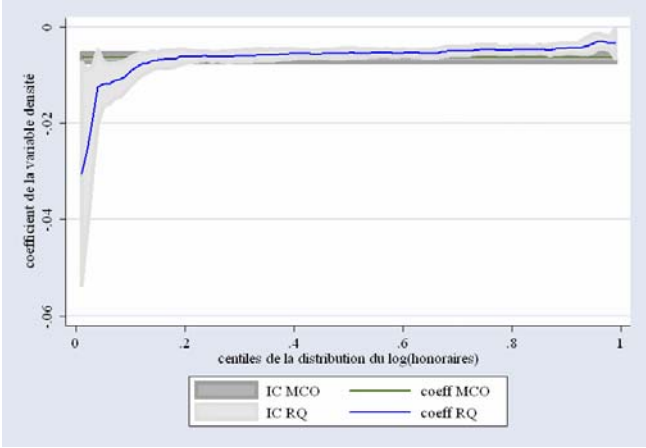
Graphique 18 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « ancienneté au carré » et intervalles de confiance à 95 % associés, année 1983



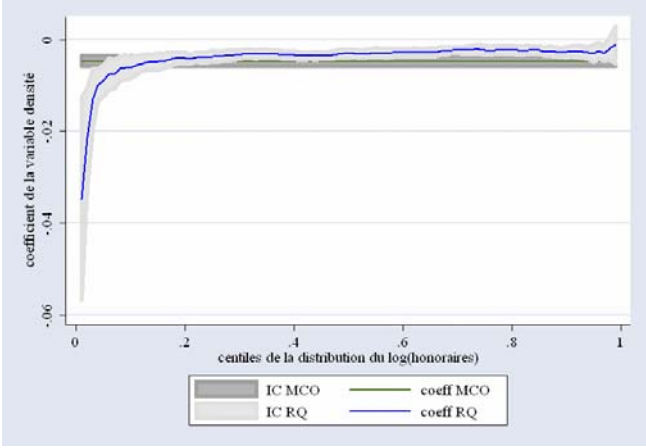
Graphique 19 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « ancienneté au carré » et intervalles de confiance à 95 % associés, année 2004



Graphique 20 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « densité médicale » et intervalles de confiance à 95 % associés, année 1983



Graphique 21 : Comparaison des coefficients obtenus par régressions quantiles ($\theta \in [0,1]$) et moindres carrés ordinaires pour la variable « densité médicale » et intervalles de confiance à 95 % associés, année 2004



2. Les résultats des régressions quantiles effectuées par sexe

Tableau 6 : Régressions quantiles et par moindres carrés ordinaires du logarithme des honoraires, pour les hommes

	1983						1992					
	MCO	0,1	0,25	0,5	0,75	0,9	MCO	0,1	0,25	0,5	0,75	0,9
Constante (écart-type)	11,638 * (0,056)	11,192 * (0,151)	11,248 * (0,057)	11,609 * (0,058)	11,838 * (0,055)	12,008 * (0,0621)	11,585 * (0,055)	11,112 * (0,128)	11,281 * (0,073)	11,593 * (0,055)	11,900 * (0,053)	12,008 * (0,059)
Ancienneté (écart-type)	0,045 * (0,0044)	0,066 * (0,008)	0,0643 * (0,005)	0,0522 * (0,0047)	0,0432 * (0,0036)	0,040 * (0,0058)	0,0447 * (0,004)	0,066 * (0,011)	0,052 * (0,004)	0,0419 * (0,004)	0,033 * (0,0045)	0,0291 * (0,004)
Ancienneté ² (écart-type)	-0,0012 * (0,0001)	-0,0018 * (0,0002)	-0,0017 * (0,0001)	-0,0013 * (0,0001)	-0,0011 * (0,0001)	-0,0010 * (0,0001)	-0,0011 * (0,0001)	-0,0018 * (0,0003)	-0,0013 * (0,0001)	-0,0011 * (0,0001)	-0,0008 * (0,0001)	-0,0007 * (0,0001)
Densité (écart-type)	-0,0067 * (0,0005)	-0,0097 * (0,0016)	-0,0061 * (0,0005)	-0,0057 * (0,0006)	-0,0049 * (0,0005)	-0,0049 * (0,0005)	-0,0049 * (0,0005)	-0,007 * (0,001)	-0,005 * (0,0006)	-0,0041 * (0,0004)	-0,0041 * (0,0005)	-0,003 * (0,0005)
R ² (ou Pseudo-R ²)	0,069	0,048	0,062	0,062	0,057	0,043	0,058	0,039	0,037	0,033	0,028	0,024

	1998						2004					
	MCO	0,1	0,25	0,5	0,75	0,9	MCO	0,1	0,25	0,5	0,75	0,9
Constante (écart-type)	11,921 * (0,067)	11,417 * (0,122)	11,639 * (0,093)	11,882 * (0,049)	12,049 * (0,064)	12,276 * (0,081)	12,011 * (0,085)	11,602 * (0,195)	11,728 * (0,082)	11,932 * (0,070)	12,104 * (0,077)	12,356 * (0,085)
Ancienneté (écart-type)	0,018 * (0,0044)	0,031 * (0,009)	0,023 * (0,006)	0,0161 * (0,004)	0,016 * (0,003)	0,018 * (0,004)	0,018 * (0,0055)	0,023 * (0,011)	0,017 * (0,006)	0,0135 * (0,004)	0,0086 ** (0,005)	0,0106 * (0,0046)
Ancienneté ² (écart-type)	-0,0005 * (0,0001)	-0,0009 * (0,0003)	-0,0006 * (0,0001)	-0,0004 * (0,0001)	-0,0004 * (0,0001)	-0,0004 * (0,0001)	-0,0005 * (0,0001)	-0,0007 * (0,0003)	-0,0004 * (0,0001)	-0,0004 * (0,0001)	-0,0002 ** (0,0001)	-0,0002 * (0,0001)
Densité (écart-type)	-0,005 * (0,0005)	-0,0064 * (0,0009)	-0,0049 * (0,0007)	-0,004 * (0,0004)	-0,0032 * (0,0005)	-0,0035 * (0,0006)	-0,0047 * (0,0006)	-0,0061 * (0,0016)	-0,0039 * (0,0006)	-0,0029 * (0,0005)	-0,002 * (0,0006)	-0,0026 * (0,0007)
R ² (ou Pseudo-R ²)	0,024	0,017	0,015	0,014	0,011	0,015	0,017	0,012	0,01	0,008	0,003	0,007

Lecture :

- « * » signifie que les coefficients sont significatifs au seuil de 5 %, et « ** » au seuil de 10 %.

Tableau 7 : Régressions quantiles et par moindres carrés ordinaires du logarithme des honoraires, pour les femmes

	1983						1992					
	MCO	0,1	0,25	0,5	0,75	0,9	MCO	0,1	0,25	0,5	0,75	0,9
Constante (écart-type)	10,931 * (0,177)	10,033 * (0,412)	10,487 * (0,198)	10,818 * (0,176)	11,230 * (0,116)	11,712 * (0,262)	10,822 * (0,145)	10,441 * (0,438)	10,657 * (0,196)	11,004 * (0,143)	11,413 * (0,118)	11,535 * (0,166)
Ancienneté (écart-type)	0,049 * (0,017)	0,130 * (0,047)	0,089 * (0,029)	0,052 ** (0,023)	0,061 * (0,015)	0,0627 * (0,019)	0,082 * (0,016)	0,138 * (0,0442)	0,0843 * (0,017)	0,057 * (0,011)	0,048 * (0,011)	0,055 * (0,012)
Ancienneté ² (écart-type)	-0,0018 * (0,0005)	-0,006 * (0,002)	-0,003 * (0,0012)	-0,0015 * (0,008)	-0,0017 * (0,0005)	-0,0017 * (0,0006)	-0,0028 * (0,0004)	-0,0055 * (0,0021)	-0,0028 * (0,0007)	-0,0018 * (0,0004)	-0,0017 * (0,0004)	-0,0017 * (0,0004)
Densité (écart-type)	-0,0048 * (0,0017)	-0,0057 (0,005)	-0,005 * (0,002)	-0,0031 * (0,0013)	-0,0040 * (0,001)	-0,0063 * (0,0026)	-0,0034 * (0,0012)	-0,009 * (0,003)	-0,0047 * (0,0017)	-0,0032 * (0,0012)	-0,0033 * (0,001)	-0,0023 (0,0015)
R ² (ou Pseudo-R ²)	0,036	0,057	0,033	0,026	0,040	0,037	0,056	0,051	0,029	0,028	0,024	0,028

	1998						2004					
	MCO	0,1	0,25	0,5	0,75	0,9	MCO	0,1	0,25	0,5	0,75	0,9
Constante (écart-type)	11,413 * (0,0083)	11,166 * (0,285)	11,124 * (0,183)	11,411 * (0,116)	11,649 * (0,124)	12,017 * (0,144)	11,655 * (0,156)	11,090 * (0,268)	11,256 * (0,172)	11,649 * (0,107)	11,810 * (0,126)	12,105 * (0,125)
Ancienneté (écart-type)	0,0296 * (0,0083)	0,027 (0,021)	0,027 * (0,0135)	0,025 * (0,008)	0,019 * (0,0097)	0,016 (0,012)	0,024 * (0,0105)	0,045 * (0,017)	0,0176 (0,012)	0,014 ** (0,009)	0,019 * (0,008)	0,012 (0,009)
Ancienneté ² (écart-type)	-0,0009 * (0,0003)	-0,001 (0,0008)	-0,0009 ** (0,0005)	-0,0008 * (0,0003)	-0,0006 ** (0,0003)	-0,0002 (0,0005)	-0,0008 * (0,0003)	-0,0017 * (0,0006)	-0,0006 (0,0004)	-0,0004 ** (0,0003)	-0,0005 ** (0,0003)	-0,0002 (0,0003)
Densité (écart-type)	-0,0042 * (0,0009)	-0,0075 * (0,002)	-0,0039 * (0,0015)	-0,0031 * (0,0009)	-0,0026 * (0,001)	-0,0038 * (0,0012)	-0,0047 * (0,0013)	-0,005 * (0,002)	-0,0025 ** (0,0013)	-0,003 * (0,0009)	-0,0027 * (0,0009)	-0,0031 * (0,001)
R ² (ou Pseudo-R ²)	0,027	0,024	0,011	0,011	0,01	0,018	0,014	0,017	0,006	0,007	0,011	0,013

Lecture :

- « * » signifie que les coefficients sont significatifs au seuil de 5 %, et « ** » au seuil de 10 %

3. Le calcul des rendements de l'ancienneté dans l'exercice de la médecine :

Le $\theta^{\text{ème}}$ quantile de la distribution conditionnelle des honoraires est défini ainsi :
 $Quant_{\theta}(y|x) = \alpha(\theta) + \beta_1(\theta) \text{sexe} + \beta_2(\theta) \text{anc} + \beta_3(\theta) \text{anc}^2 + \beta_4(\theta) \text{dens}$.

Sa dérivée par rapport à l'ancienneté, qui permet de mesurer le rendement de l'ancienneté, est donc $\frac{\partial Q_y(\theta|x)}{\partial(\text{anc})} = r = \beta_2(\theta) + 2\beta_3(\theta) \text{anc}$: elle tient compte de l'influence quadratique de

l'ancienneté sur le logarithme des honoraires.

Ce rendement dépend donc de la valeur de l'ancienneté à laquelle on l'évalue.

Il diffère alors :

- en fonction du quantile auquel la relation est estimée
- en fonction du nombre d'années d'ancienneté déjà acquis

Les tableaux suivants calculent les rendements pour l'ensemble des régressions quantiles estimées, à savoir hommes et femmes confondus, mais aussi séparément.

Tableau 8: Calcul des rendements de l'ancienneté correspondant aux régressions quantiles estimées
hommes
et femmes confondus :

	1983				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,056	0,036	0,016	0,002	-0,020
Q25	0,058	0,038	0,018	0,005	-0,018
Q50	0,046	0,033	0,020	0,006	-0,004
Q75	0,041	0,030	0,019	0,005	-0,001
Q90	0,038	0,028	0,018	0,005	0,000
MCO	0,039	0,027	0,015	0,0034	-0,006

	1992				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,067	0,047	0,027	0,003	-0,009
	0,049	0,035	0,021	0,006	-0,004
	0,039	0,028	0,017	0,004	-0,003
	0,030	0,021	0,012	0,003	-0,004
	0,028	0,021	0,014	0,005	0,002
	0,044	0,034	0,024	0,005	0,006

	1998				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,028	0,019	0,010	-0,0006	-0,006
Q25	0,019	0,014	0,009	0,002	0,0000
Q50	0,014	0,010	0,006	0,003	-0,001
Q75	0,013	0,010	0,006	0,002	-0,001
Q90	0,016	0,012	0,008	0,003	0,001
MCO	0,018	0,013	0,008	0,001	-0,001

	2004				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,019	0,012	0,0049	-0,002	-0,008
	0,015	0,011	0,0074	0,0036	0,0005
	0,011	0,008	0,0039	0,0000	-0,003
	0,011	0,008	0,0049	0,0014	-0,001
	0,013	0,009	0,006	0,003	0,0003
	0,016	0,010	0,005	-0,0003	-0,005

Tableau 9 : Calcul des rendements de l'ancienneté correspondant aux régressions quantiles estimées
pour les hommes

	1983				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,059	0,040	0,022	0,003	-0,012
Q25	0,057	0,040	0,023	0,006	-0,007
Q50	0,047	0,033	0,020	0,006	-0,005
Q75	0,039	0,028	0,017	0,006	-0,003
Q90	0,036	0,026	0,016	0,005	-0,003
MCO	0,040	0,028	0,016	0,004	-0,005

	1992				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,059	0,041	0,023	0,006	-0,008
	0,046	0,033	0,019	0,006	-0,004
	0,037	0,027	0,016	0,005	-0,004
	0,029	0,021	0,013	0,005	-0,002
	0,026	0,019	0,012	0,005	-0,0006
	0,040	0,029	0,017	0,006	-0,003

	1998				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,028	0,019	0,010	0,0007	-0,006
Q25	0,021	0,015	0,009	0,003	-0,001
Q50	0,015	0,011	0,007	0,003	0,000
Q75	0,014	0,010	0,006	0,003	0,000
Q90	0,016	0,012	0,007	0,003	-0,001
MCO	0,016	0,011	0,006	0,001	-0,002

	2004				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,020	0,013	0,006	-0,0008	-0,006
	0,015	0,010	0,006	0,001	-0,002
	0,012	0,008	0,004	0,0002	-0,003
	0,007	0,005	0,003	0,0006	-0,001
	0,009	0,007	0,004	0,002	-0,0003
	0,016	0,011	0,0053	0,0014	-0,0040

Tableau 10 : Calcul des rendements de l'ancienneté correspondant aux régressions quantiles estimées pour les femmes

	1983				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,107	0,048	-0,010	-0,069	-0,116
Q25	0,077	0,047	0,016	-0,014	-0,038
Q50	0,046	0,030	0,015	0,0002	-0,012
Q75	0,054	0,037	0,020	0,003	-0,011
Q90	0,056	0,039	0,022	0,005	-0,008
MCO	0,042	0,024	0,006	-0,011	-0,025

	1992				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,116	0,061	0,006	-0,049	-0,094
	0,073	0,044	0,015	-0,013	-0,036
	0,049	0,031	0,013	-0,005	-0,019
	0,041	0,025	0,008	-0,008	-0,022
	0,048	0,030	0,013	-0,005	-0,018
	0,070	0,042	0,014	-0,014	-0,036

	1998				
	2 ans	7 ans	12 ans	17 ans	21 ans
Q10	0,023	0,012	0,001	-0,009	-0,018
Q25	0,023	0,014	0,005	-0,005	-0,012
Q50	0,021	0,013	0,005	-0,003	-0,010
Q75	0,017	0,011	0,004	-0,002	-0,007
Q90	0,015	0,013	0,010	0,007	0,006
MCO	0,026	0,016	0,006	-0,003	-0,011

	2004				
	2 ans	7 ans	12 ans	17 ans	21 ans
	0,038	0,021	0,003	-0,014	-0,028
	0,015	0,009	0,003	-0,002	-0,007
	0,013	0,008	0,004	-0,0004	-0,004
	0,017	0,012	0,007	0,003	-0,0008
	0,011	0,009	0,007	0,005	0,004
	0,021	0,013	0,005	-0,003	-0,009

4. Les résultats des régressions interquantiles

Les régressions quantiles n'apportent une valeur ajoutée à l'estimation par moindres carrés ordinaires que si les coefficients estimés des différentes régressions quantiles sont significativement différents les uns des autres. Si ce n'est pas le cas, le coefficient unique obtenu par moindres carrés ordinaires résume suffisamment bien l'information sur la relation entre les variables explicatives et la variable expliquée. C'est le *test de Wald* d'égalité des coefficients (annexe 2) qui permet de tester si les coefficients estimés diffèrent significativement ou non entre les quantiles.

Une autre méthode, mais équivalente, consiste à mettre en œuvre des *régressions interquantiles*.

En effet, les régressions simultanées du 1er et du 3ème quartile estiment l'influence des variables explicatives au 1er et au 3ème quartile de la distribution conditionnelle de la variable expliquée Y .

Les deux estimations obtenues sont alors : $\hat{y}_{q_1} = \hat{\alpha}_{q_1} + \hat{\beta}_{q_1}x$ et $\hat{y}_{q_3} = \hat{\alpha}_{q_3} + \hat{\beta}_{q_3}x$.

Pour tester si l'influence de la variable x est la même au 1er et au 3ème quartile de la distribution conditionnelle de Y , le test de Wald teste l'hypothèse :

$$H_0 : \hat{\beta}_{q_1} = \hat{\beta}_{q_3} \text{ contre } H_A : \hat{\beta}_{q_1} \neq \hat{\beta}_{q_3} \text{ }^{36}$$

³⁶ Il est également possible de tester si la valeur de la constante est significativement différente entre les quantiles ; le test s'effectue de manière identique. Toutefois, à la différence des différentes variables explicatives d'un modèle dont l'influence peut être, dans certains cas, identique à tous les quantiles de la distribution conditionnelle, il est plus que probable que la constante du 3^{ème} quartile soit significativement différente de celle du 1^{er} quartile.

Le résultat du test nous indique si le coefficient associé à la variable x varie significativement entre les quantiles.

Parallèlement, il est possible de mettre en œuvre des régressions interquantiles. Les plus courantes sont les régressions *interquartiles* et *interdéciles*.

La régression interquartile consiste ainsi à estimer l'écart entre le 1er et le 3ème quartile de la distribution de Y , conditionnellement aux valeurs de X .

L'estimation obtenue directement est alors : $\hat{y}_{(q3-q1)} = \hat{\alpha}_{(q3-q1)} + \hat{\beta}_{(q3-q1)}x$

Mais effectuer cette régression interquartile (ou interquantile dans le cas plus général) revient :

- À estimer les régressions du 1er quartile et du 3ème quartile séparément, soit :
 $\hat{y}_{q1} = \hat{\alpha}_{q1} + \hat{\beta}_{q1}x$ et $\hat{y}_{q3} = \hat{\alpha}_{q3} + \hat{\beta}_{q3}x$
- Puis à soustraire ces deux expressions. Ainsi, $\hat{y}_{(q3-q1)} = \underbrace{(\hat{\alpha}_{q3} - \hat{\alpha}_{q1})}_{\hat{\alpha}_{(q3-q1)}} + \underbrace{(\hat{\beta}_{q3} - \hat{\beta}_{q1})}_{\hat{\beta}_{(q3-q1)}}x$

Dès lors, si $\hat{\beta}_{(q3-q1)}$ est significativement différent de zéro, l'influence de la variable x est significativement différente entre le 1er et le 3^{ème} quartile : elle n'est donc pas la même à tous les points de la distribution conditionnelle de Y . Cela revient bien à effectuer le test de Wald énoncé plus haut.

Tableau 11 : Régressions interquartiles et interdéciles du logarithme des honoraires, hommes et femmes confondus

	1983						1992					
	75/25	50/25	75/50	90/10	50/10	90/50	75/25	50/25	75/50	90/10	50/10	90/50
Cons. (e.t)	0,375* (0,080)	0,260* (0,069)	0,115** (0,059)	0,471* (0,202)	0,254 (0,188)	0,217* (0,088)	0,47* (0,081)	0,248* (0,069)	0,221* (0,054)	0,623* (0,158)	0,355* (0,143)	0,268* (0,075)
Sexe (e.t)	0,181* (0,042)	0,069** (0,037)	0,113* (0,032)	0,363* (0,107)	0,157** (0,094)	0,206* (0,051)	0,156* (0,031)	0,075* (0,028)	0,081* (0,021)	0,324* (0,058)	0,182* (0,055)	0,142* (0,031)
Anc (e.t)	-0,021* (0,005)	-0,014* (0,005)	-0,007** (0,004)	-0,022* (0,009)	-0,012 (0,008)	-0,009** (0,006)	-0,021* (0,005)	-0,011* (0,004)	-0,009* (0,004)	-0,044* (0,010)	-0,032* (0,009)	-0,012* (0,004)
Anc ² (e.t)	0,0006* (0,0001)	0,0004* (0,0001)	0,0002** (0,0001)	0,0007* (0,0003)	0,0005* (0,0002)	0,0003** (0,0001)	0,0005* (0,0001)	0,0003* (0,0001)	0,0002* (0,0001)	0,0013* (0,0003)	0,0009* (0,0003)	0,0004* (0,0001)
Dens (e.t)	0,0014* (0,0006)	0,0008** (0,0005)	0,0006 (0,0004)	0,0049* (0,0016)	0,004* (0,002)	0,001 (0,0006)	0,0007 (0,0006)	0,0006 (0,0005)	0,00009 (0,0004)	0,0043* (0,001)	0,003* (0,001)	0,0012* (0,0005)

	1998						2004					
	75/25	50/25	75/50	90/10	50/10	90/50	75/25	50/25	75/50	90/10	50/10	90/50
Cons. (e.t)	0,286* (0,092)	0,163* (0,078)	0,123* (0,054)	0,683* (0,141)	0,387* (0,117)	0,295* (0,080)	0,335* (0,089)	0,217* (0,071)	0,118* (0,06)	0,551* (0,170)	0,243* (0,160)	0,307* (0,072)
Sexe (e.t)	0,103* (0,028)	0,062* (0,024)	0,041* (0,017)	0,201* (0,041)	0,112* (0,033)	0,089* (0,023)	0,064* (0,026)	0,026 (0,022)	0,038* (0,016)	0,152* (0,047)	0,080 (0,042)	0,072* (0,02)
Anc (e.t)	-0,006 (0,005)	-0,0053 (0,0046)	-0,0007 (0,003)	-0,013 (0,009)	-0,016* (0,008)	0,003 (0,004)	-0,033 (0,005)	-0,0035 (0,0042)	0,0002 (0,004)	-0,008 (0,009)	-0,009 (0,009)	0,0008 (0,004)
Anc ² (e.t)	0,0002 (0,0001)	0,0002 (0,0001)	8,38e-06 (0,00001)	0,0005** (0,0003)	0,0006* (0,0002)	-0,00005 (0,0001)	0,0001 (0,0001)	0,00008 (0,0001)	0,00003 (0,0001)	0,0004 (0,0002)	0,0003* (0,0002)	0,00006 (0,0001)
Dens (e.t)	0,0016* (0,0007)	0,0009 (0,0006)	0,0008* (0,0004)	0,0024* (0,001)	0,002* (0,0008)	0,0004 (0,0006)	0,0014* (0,0006)	0,0006 (0,0005)	0,0007** (0,0004)	0,003* (0,001)	0,003 (0,001)	0,0003 (0,0005)

Lecture :

- « * » signifie que les coefficients sont significatifs au seuil de 5 %, et « ** » au seuil de 10 %
- En 1983, le coefficient associé à la constante pour la régression interquartile est 0,375.
- Cela signifie que la constante de la régression du troisième quartile est de 0,375 point plus élevée que celle du premier quartile.
- En effet, la valeur de la constante pour RQ(0,25)=11,891 et pour RQ(0,75)=12,267 (cf. tableau 2). La différence des deux coefficients est donc bien 0,375.
- Par ailleurs, ce coefficient est significativement différent de zéro : la constante n'est pas la même à tous les quantiles de la distribution du logarithme des honoraires.

VI. Annexe 2 : La méthode des régressions quantiles

1. Introduction

Développées par Koenker et Bassett dès 1978, les régressions quantiles ont constitué une alternative aux moindres carrés ordinaires pour l'estimation de modèles économétriques.

Alors que les moindres carrés ordinaires décrivent comment *la moyenne conditionnelle* d'une variable Y dépend des variables explicatives X , les régressions quantiles indiquent comment les différents quantiles de la distribution conditionnelle de Y dépendent des variables explicatives. Chaque régression quantile caractérisant un point particulier de la distribution conditionnelle, l'estimation de plusieurs régressions quantiles en procure une description complète.

Prenons l'exemple de la mesure du rendement de l'expérience grâce à l'estimation d'une fonction de gains. Les régressions quantiles permettent en particulier de répondre à deux questions :

- Comment varie généralement le profil des salaires avec l'expérience ?
- Quelle est l'influence de l'expérience sur les individus ayant les salaires les plus bas et les plus élevés ?

La mise en œuvre de la régression médiane donne une réponse à la première question, les moindres carrés ordinaires étant moins robustes que cette dernière pour étudier des distributions étalées (ce qui est généralement le cas de la distribution des salaires).

Par ailleurs, si on considère comme « pauvres » les individus se situant au-dessous du premier décile de la distribution des salaires et comme « riches » ceux qui sont au-dessus du neuvième décile, c'est la mise en œuvre de la 10^{ème} et de la 90^{ème} régression quantile qui apporte une réponse à la deuxième question.

Plus généralement, il semble improbable que l'ensemble des variables explicatives (nombre d'années d'études, expérience sur le marché du travail, sexe, etc.) ait un impact identique en tous points de la distribution des salaires.

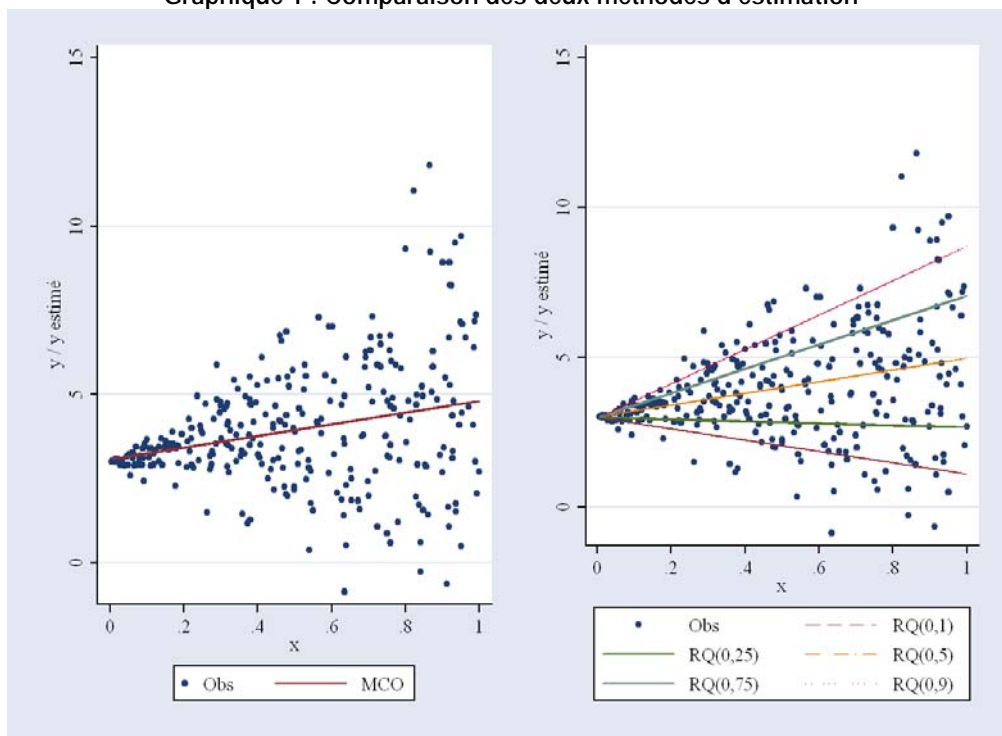
En conséquence, les régressions quantiles, qui permettent d'estimer l'ensemble d'une distribution conditionnelle sont un outil efficace pour les études économétriques s'intéressant à des caractéristiques de dispersion de la variable expliquée.

Un exemple graphique permet de comprendre la différence entre l'estimation par moindres carrés ordinaires et par la méthode des régressions quantiles. Pour cela, un échantillon de trois cent observations est simulé, tel que $X \rightarrow U(0;10)$, $u \rightarrow N(0;9x^2)$ et $y = 3 + 2x + u$.

La partie gauche graphique 1 représente la droite estimée par les moindres carrés ordinaires. La partie droite représente, elle, cinq droites estimées, correspondant à l'estimation de la dixième, vingt-cinquième, cinquantième, soixante-quinzième et quatre-vingt-dixième régression quantile. Alors que dans le cas des moindres carrés ordinaires, une seule droite

résume toute la relation entre les variables X et Y , ce sont ici cinq droites qui font de même dans le cas de l'estimation par régressions quantiles³⁷.

Graphique 1 : Comparaison des deux méthodes d'estimation



Cette annexe s'attache à décrire la méthode des régressions quantiles, en étudiant ses origines, ses hypothèses, sa mise en œuvre et ses propriétés.

2. Les prémices des régressions quantiles

A. Hogg (1975) et les droites de pourcentage

Si la méthode des régressions quantiles a été développée par Koenker et Bassett (1978), Hogg (1975) en avait déjà donné les intuitions : selon lui, l'estimation de plusieurs droites de pourcentage plutôt qu'une droite de régression moyenne résume mieux la relation entre deux ou plusieurs variables.

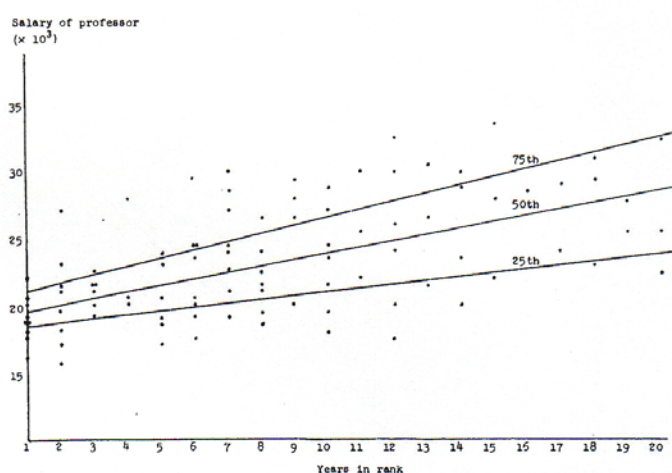
Dans son article, il souhaite estimer la relation entre les salaires de quatre-vingt seize professeurs de statistiques (variable y) et leur ancienneté, inférieure ou égale à vingt ans (variable x), mais en utilisant une autre méthode d'estimation que les moindres carrés ordinaires. En effet, le graphique 2, qui décrit son échantillon d'observations, indique une dispersion des observations croissante avec l'ancienneté ainsi que la présence de valeurs atypiques (notamment pour de faibles valeurs du salaire), éléments qui rendent la méthode des moindres carrés ordinaires peu robuste.

³⁷ Par souci de lisibilité, seules cinq régressions quantiles, notées $RQ(\theta)$ ont été mises en œuvre (RQ(0,10), RQ(0,25), RQ(0,50), RQ(0,75), RQ(0,90)). En pratique, il est possible d'en estimer beaucoup d'autres, pour tout $\theta \in [0,1]$.

Son approche est essentiellement graphique et manuelle. Il se place dans le repère (ancienneté, salaire) et sépare les observations en deux groupes de taille identique, en fonction de la valeur de x : les professeurs dont l'ancienneté est inférieure ou égale à la médiane³⁸ (sept ans) appartiennent au premier groupe, les autres au second.

Pour trouver l'estimation du $\theta^{\text{ème}}$ quantile de la distribution des salaires, il trace une droite séparant ses observations, de telle sorte que $(1-\theta)\%$ des observations de chacun des deux groupes précédents soient au-dessus de cette droite et $\theta\%$ en dessous. Il obtient alors l'estimation de la droite de régression du $\theta^{\text{ème}}$ quantile. Le graphique 2 illustre sa démarche, pour les droites de régression des 25ème, 50ème et 75ème quantiles.

Graphique 2 : L'estimation des « droites de pourcentage » (Hogg)
The Case of 96 Professors' Salaries



Source : Hogg (1975)

B. La recherche d'alternatives aux Moindres carrés ordinaires

En 1978, la méthode des moindres carrés ordinaires était la méthode d'estimation la plus courante, mais elle était également très critiquée pour trois de ses « faiblesses »³⁹.

Tout d'abord, le modèle de régression linéaire classique résume l'influence de variables explicatives X sur l'ensemble de la distribution de la variable expliquée y à l'aide d'un seul coefficient estimé. Or, cette unique mesure peut ne pas être suffisante pour synthétiser toute la relation entre ces variables. Par ailleurs, d'autres mesures peuvent être plus révélatrices de la tendance d'un échantillon que la moyenne, comme la médiane par exemple.

³⁸ Le choix de la médiane de X est un moyen d'avoir autant d'observations dans chacun des deux groupes, pour rendre l'estimation plus précise. Hogg généralise cependant sa méthode à l'utilisation d'un découpage différent : si $V(y|x \text{ faible}) < V(y|x \text{ élevé})$, l'estimation nécessitera moins d'observations dans le premier groupe que dans le second, et c'est une valeur de X inférieure à la médiane qui délimitera alors les deux groupes.

³⁹ La méthode des moindres carrés ordinaires aurait été utilisée pour la première fois par A.-M. Legendre en 1805, dans des travaux relatifs à l'astronomie (Cf. Dormont, 1999).

Ensuite, les estimations par moindres carrés ordinaires sont très sensibles aux valeurs atypiques des variables expliquées et explicatives, qui peuvent influencer la valeur des coefficients estimés.

Par ailleurs, l'hypothèse de normalité des perturbations qui sous-tend les estimations par moindres carrés ordinaires⁴⁰ est rarement vérifiée en pratique.

« Everyone believes in the Gaussian law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact » (Poincaré, cité par Koenker et Bassett (1978)).

En pratique, les distributions d'erreurs ont souvent des queues plus larges que celles attendues avec la distribution gaussienne, ce qui rend les moindres carrés ordinaires moins précis. Dans ce cas, d'autres méthodes d'estimations peuvent se révéler plus efficaces que les moindres carrés ordinaires, et c'est notamment le cas des régressions quantiles.

Les recherches se sont d'abord concentrées sur le « Location Model »⁴¹ pour mettre en œuvre une théorie statistique basée sur les quantiles d'un échantillon. La recherche d'estimateurs robustes sur ces modèles a conduit au développement de plusieurs méthodes économétriques, les estimateurs de classe « L » étant les plus connus.

- Parmi eux, les « Least Trimmed Squares » ou le « α -trimmed mean » consistent à estimer la moyenne de l'échantillon après avoir supprimé les $\alpha\%$ d'observations les plus élevées et les plus faibles. C'est un moyen d'éliminer les valeurs atypiques.
- Laplace, dès 1818, propose l'estimateur des moindres écarts absolus (Least Absolute Estimator, estimateur LAE). Ceui-ci consiste à minimiser $|u_i|$ plutôt que $(u_i)^2$ et à estimer, par conséquent, la médiane conditionnelle de la variable dépendante plutôt que la moyenne.

Koenker et Bassett (1978) ont développé les « régressions quantiles » pour étendre l'estimation des quantiles dans le cadre du « location model » à une classe de modèles linéaires plus générale : les quantiles conditionnels ayant une forme linéaire.

⁴⁰ L'estimation par MCO suppose que les perturbations sont indépendantes et identiquement distribuées, mais aussi, même si cette condition n'est pas toujours imposée, qu'elles suivent une loi Normale. Cette dernière hypothèse n'est pas nécessaire pour obtenir la plupart des résultats du modèle linéaire multiple, mais lorsqu'elle est spécifiée, la méthode des MCO fournit l'estimation du *maximum de vraisemblance* de la moyenne conditionnelle. Par ailleurs, elle est utile pour la construction de tests statistiques et d'intervalles de confiance associés aux coefficients estimés (Cf. Greene (2003), pp.17). Par exemple, la distribution de la statistique t (tests de significativité sur la valeur d'un coefficient) et de la statistique f (tests de contraintes linéaires sur les coefficients) reposent sur l'hypothèse de normalité des perturbations.

Lorsque cette hypothèse n'est pas vérifiée, l'estimateur des MCO ne suit pas une loi normale à distance finie et il faut se placer dans le cadre où il est asymptotiquement normal. Les tests sont alors basés sur la normalité asymptotique des MCO mais ils nécessitent toutefois un très grand nombre d'observations afin que la convergence en loi des différentes statistiques soit réalisée (Cf. Dormont (1999), pp.241-250).

⁴¹ Modèle du type $y_i = \beta + u_i$ (c'est à dire avec une seule variable explicative x_i qui vaut 1 pour $i=1\dots n$).

3. Rappels sur les fonctions quantiles

A. Aperçu des principales notations utilisées dans les différentes sections

Lien entre $a(\theta)$ et $Q_Y(\theta)$

- Nous noterons $a(\theta)$ le $\theta^{\text{ème}}$ quantile non conditionnel d'une variable aléatoire ou d'un échantillon. Toutefois, ce quantile peut ne pas être unique (voir section 3.2).
 - Par leur définition même (voir section 3.3), la fonction quantile, $Q_Y(\theta)$, et la fonction quantile empirique, $\hat{Q}_Y(\theta)$, permettent d'obtenir un unique $\theta^{\text{ème}}$ quantile d'une variable aléatoire ou d'un échantillon.
- ⇒ $Q_Y(\theta)$ ou $\hat{Q}_Y(\theta)$ donnent le plus petit élément de l'ensemble des $a(\theta)$ obtenus par ordonnancement des observations ou optimisation (voir section 4).

Les quantiles non conditionnels

- Soit (y_1, \dots, y_n) un échantillon. $Q_Y(\theta)$ est le $\theta^{\text{ème}}$ quantile non conditionnel de cet échantillon si c'est la plus petite valeur y_i de l'échantillon telle que $\theta\%$ des observations lui sont inférieures (voir section 3.3).

Les quantiles non conditionnels sont les statistiques d'ordre d'un échantillon. Ils s'obtiennent par simple ordonnancement des observations.

Les quantiles conditionnels

- Soit $(x_i, y_i)_{i=1, \dots, n}$ un échantillon. On suppose que les variables y_i sont fonction des variables x_i et on note $Q_Y(\theta|x)$ le $\theta^{\text{ème}}$ quantile conditionnel. L'ordonnancement des observations ne permet plus d'obtenir ces quantiles directement.
- Il faut faire ici un parallèle avec le cas de la moyenne μ d'un échantillon. Si on suppose une relation linéaire entre y et x , alors la régression par moindres carrés ordinaires permet d'obtenir la moyenne conditionnelle de y sachant x , notée $\mu(x) = x' \beta$.
- De la même façon, ce sont les régressions quantiles qui permettent d'obtenir les quantiles conditionnels. Ainsi, et toujours en supposant une relation linéaire entre les variables y et x , $Q_Y(\theta|x) = x' \beta(\theta)$ est le $\theta^{\text{ème}}$ quantile conditionnel de y sachant x . Il est défini tel que $\theta\%$ des observations sont telles que $y_i \leq x_i \beta(\theta)$ et $(1 - \theta)\%$ des observations sont telles que $y_i > x_i \beta(\theta)$.

B. Définition du $\theta^{\text{ème}}$ quantile non conditionnel d'une variable aléatoire

Soit Y une variable aléatoire (discrète ou continue) à valeurs dans \mathfrak{R} .

Pour $\theta \in [0,1]$, le $\theta^{\text{ème}}$ quantile de la distribution d'une variable aléatoire Y est défini comme la (ou les) valeur(s) $a(\theta)$ telle que :

- $\Pr(Y \leq a(\theta)) \geq \theta$
- $\Pr(Y \geq a(\theta)) \geq 1 - \theta$

Ainsi, $\Pr(Y < a(\theta)) \leq \theta \leq \Pr(Y \leq a(\theta))$

C'est donc la (ou les) valeur(s) $a(\theta)$ telle que, pour une proportion θ de la population, Y prend une valeur inférieure ou égale à $a(\theta)$.

Par exemple, un omnipratricien au dixième quantile de la distribution non conditionnelle des honoraires a une rémunération supérieure à celle des 10 % d'omnipratriciens les moins bien rémunérés, mais inférieure à celle des 90 % restants.

Remarque 1

On appelle ici « variable continue » une variable qui prend ses valeurs dans un intervalle. Les « variables discrètes » ne prennent qu'un nombre fini de valeurs.

Remarque 2

Si Y est une variable aléatoire continue, on aura, dans la plupart des cas : $\Pr(Y < a) = \Pr(Y \leq a) = \theta$. Le quantile d'ordre θ est donc le nombre $a(\theta)$ tel que $\Pr(Y \leq a(\theta)) = \theta$

Si Y est une variable aléatoire discrète, on cherche le plus petit y_i dans le support de Y pour lequel $F(y_i) \geq \theta$.

- Si $F(y_i) > \theta$, alors $a(\theta)$ est unique et est égal à ce y_i .
- Mais si $F(y_i) = \theta$, alors tout nombre réel dans l'intervalle $[y_i, y_{i+1}]$ correspond à un quantile d'ordre θ .

Définissons maintenant la « **fonction quantile** », notée $Q_Y(\theta)$: elle assure une solution unique pour le $\theta^{\text{ème}}$ quantile de Y , le plus petit élément de l'intervalle de solutions décrit précédemment.

C. Lien entre la fonction de répartition et la fonction quantile d'une variable aléatoire

Soit Y une variable aléatoire à valeurs dans \mathfrak{R} et $F_Y(y) = \Pr(Y \leq y)$ sa fonction de répartition.

$$\boxed{\begin{array}{l} F_Y : \mathfrak{R} \rightarrow [0,1] \\ y \mapsto F_Y(y) = \Pr(Y \leq y) \end{array}} \quad (1)$$

Cette fonction de répartition possède plusieurs propriétés :

- Elle est croissante de 0 à 1 quand y passe de $-\infty$ à $+\infty$.
- Elle vérifie $\lim_{y \rightarrow -\infty} F_Y(y) = 0$ et $\lim_{y \rightarrow +\infty} F_Y(y) = 1$
- Elle est continue à droite et admet en tout point y_0 une limite à gauche :

$$F(y_0^+) = \lim_{y \rightarrow y_0^+} F_Y(y) = F(y_0) \text{ et } F(y_0^-) = \lim_{y \rightarrow y_0^-} F_Y(y) = \Pr(Y < y_0)$$

Remarque 1

Si Y est une variable aléatoire discrète, sa fonction de répartition est une fonction en escalier. Si y_1, \dots, y_n sont les réalisations de la variable aléatoire Y , rangées par ordre croissant, alors la fonction de répartition $F_Y(y)$ prend les valeurs :

$$F_Y(y) = \begin{cases} 0 & \text{si } y < y_1 \\ P(y = y_1) & \text{si } y \in [y_1, y_2] \\ \vdots & \\ P(y = y_1) + \dots + P(y = y_k) & \text{si } y \in [y_k, y_{k+1}] \\ \vdots & \end{cases}$$

Si Y est une variable continue, sa fonction de répartition $F_Y(y)$ est la primitive de la densité de probabilité qui s'annule en $-\infty$: $F_Y(y) = \int_{-\infty}^y f(u)du$.

C'est une fonction *continue en tout point*.

En pratique, les fonctions en escalier peuvent souvent être approximées par des fonctions continues, une variable discrète se rapprochant souvent d'une variable continue, lorsqu'on multiplie les expériences aléatoires Y et affine les classes.

Remarque 2

La fonction de répartition est la fonction qui, à tout nombre réel, associe le nombre d'individus de la population étudiée qui ont une valeur inférieure à ce réel.

- Dans les pays anglo-saxons, la fonction de répartition est définie comme en (1) : par inférieure, on entend inférieure ou égale.
- En revanche, en France, la fonction de répartition est traditionnellement définie ainsi :

$F_Y : \mathfrak{R} \rightarrow [0,1]$
$y \mapsto F_Y(y) = \Pr(Y < y)$

Par inférieure, on entend *strictement inférieure*. En conséquence, c'est la continuité à gauche et non plus à droite de la fonction $F_Y(y)$ qui est posée.

Ici, nous choisissons la formulation anglo-saxonne, pour plus de cohérence avec l'ensemble des études menées sur les régressions quantiles.

Si F est continue et *strictement* croissante sur \mathfrak{R} , elle réalise une bijection de \mathfrak{R} sur $[0,1]$ et admet un inverse, noté F^{-1} .

Mais cette fonction F peut ne pas être strictement croissante (constante sur certains intervalles) ; elle n'est donc pas toujours bijective et n'admet pas toujours d'inverse. On définit alors l'inverse généralisé de F .

La fonction quantile d'une variable aléatoire Y est donc l'inverse de sa fonction de répartition ou son inverse généralisé.

$$\boxed{F_Y^{-1} : [0,1] \rightarrow \mathfrak{R}} \quad (2)$$

$$\theta \mapsto F_Y^{-1}(\theta) = \inf \{y \text{ t.q. } F_Y(y) \geq \theta\}$$

On note la fonction quantile : $Q_Y(\theta)$.

Par convention :

- $Q_Y(0)$ représente la plus petite des valeurs possibles pour Y : $Q_Y(0) = \min\{y_i\}$
- $Q_Y(1)$ représente la plus grande : $Q_Y(1) = \max\{y_i\}$.

Remarque : Les valeurs $Q_Y(0)$ et $Q_Y(1)$ peuvent être infinies.

Exemples

Prenons deux exemples qui montrent le lien entre la fonction de répartition et la fonction quantile d'une variable aléatoire.

1. Soit $F_Y(y) = (1 - e^{-\lambda y})$ la fonction de répartition de la loi exponentielle $\varepsilon(\lambda)$.

La fonction quantile $Q_Y(\theta)$ est telle que : $\forall \theta \in [0,1], (1 - e^{-\lambda y}) = \theta \Leftrightarrow y = Q_Y(\theta)$.

Alors $Q_Y(\theta) = -\frac{1}{\lambda} \log(1 - \theta)$.

2. Soit $F_Y(y) = y$ la fonction de répartition de la loi uniforme $U(0,1)$.

Alors la fonction quantile est telle que $Q_Y(\theta) = \theta$.

Tout comme la fonction de répartition, la fonction quantile permet de caractériser entièrement la distribution d'une variable aléatoire Y .

La fonction quantile possède plusieurs propriétés :

- Elle est croissante quand θ varie de 0 à 1
- Elle est continue à gauche
- Elle admet une limite à droite en tout point $\theta \in [0,1]$.

D. De la fonction quantile théorique à la fonction quantile empirique

Définition du $\theta^{\text{ème}}$ quantile non conditionnel d'un échantillon

Soit $Y = (y_1, \dots, y_n)$ un échantillon et $(y_{(1)}, \dots, y_{(n)})$ les statistiques d'ordre de l'échantillon, c'est-à-dire les y_i rangés par ordre croissant : $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$.

Le $\theta^{\text{ème}}$ quantile d'un échantillon est le (ou les) nombres $a(\theta)$ tel que :

$$\boxed{\frac{1}{n} \sum_{i=1}^n I(y_i \leq a(\theta)) \geq \theta \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^n I(y_i \geq a(\theta)) \geq 1 - \theta} \quad (3)$$

Par exemple, si $\theta = 0,75$, le $75^{\text{ème}}$ quantile est la valeur a de l'échantillon telle que 75 % des observations lui sont inférieures et 25 % lui sont supérieures.

Remarque

Soit l'échantillon (y_1, \dots, y_n) , où les différentes observations sont rangées par ordre croissant.

- Les trois quartiles d'un échantillon sont notés : $Q_1 = Q(0,25)$, $Q_2 = Q(0,5)$ et $Q_3 = Q(0,75)$
- Les neuf déciles sont les valeurs de $Q\left(\frac{i}{10}\right)$, $i = 1 \dots 9$.
- La médiane est une valeur particulière de l'échantillon. C'est à la fois le second quartile, le cinquième décile, etc.

Les fonctions de répartition et fonctions quantiles empiriques

Soit $Y = (y_1, \dots, y_n)$ un échantillon de fonction de répartition F et $(y_{(1)}, \dots, y_{(n)})$ les statistiques d'ordre de l'échantillon.

On appelle fonction de répartition empirique associée à (y_1, \dots, y_n) la fonction aléatoire $\hat{F}_Y(y)$ définie ainsi :

$$\boxed{\hat{F}_Y : \mathfrak{R} \rightarrow [0,1] \quad y \mapsto \hat{F}_Y(y) = \begin{cases} 0 & \text{si } y < y_{(1)} \\ \frac{i}{n} & \text{si } y_{(i)} \leq y < y_{(i+1)} \\ 1 & \text{si } y \geq y_{(n)} \end{cases} \quad (4)}$$

$\hat{F}_Y(y)$ représente donc la proportion d'éléments de l'échantillon qui sont inférieurs ou égaux à y .

En notant $I(P)$ la fonction indicatrice qui vaut 1 si l'expression P est vraie, la fonction de répartition empirique s'exprime également ainsi : $\hat{F}_Y(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$.

Remarque

Dans le cas d'un échantillon discret, la fonction de répartition empirique est représentée par un diagramme en bâtons.

Dans le cas d'un échantillon continu, la fonction de répartition empirique est représentée par un histogramme.

La fonction quantile empirique de ce même échantillon, $(y_{(1)}, \dots, y_{(n)})$, notée $\hat{Q}_Y(\theta)$, est la fonction quantile de sa distribution empirique. Elle est définie ainsi :

$$\hat{F}_Y^{-1} : [0,1] \rightarrow \mathfrak{R}$$

$$\theta \mapsto \hat{Q}_Y(\theta) = \begin{cases} 0 & \text{si } y < y_{(1)} \\ y_{(i)} & \text{si } \frac{i-1}{n} < \theta \leq \frac{i}{n} \\ 1 & \text{si } y \geq y_{(n)} \end{cases} \quad (5)$$

Si la série est rangée par ordre croissant, le $\theta^{\text{ème}}$ quantile $\hat{Q}_Y(\theta)$ est la valeur du terme de cette série dont l'indice (i) est le plus petit entier supérieur ou égal à $n\theta$. Autrement dit, il y a $n\theta$ observations sous le $\theta^{\text{ème}}$ quantile $\hat{Q}_Y(\theta)$ et $n(1-\theta)$ au niveau ou au-dessus de ce quantile.

Cette fonction peut également être notée :

$$\hat{F}_Y^{-1} : [0,1] \rightarrow \mathfrak{R}$$

$$\theta \mapsto \hat{Q}_Y(\theta) = \inf \{y_i \text{ t.q. } \hat{F}_Y(y_i) \geq \theta\}$$

Remarque 1

Si la distribution d'un échantillon est symétrique, la moyenne et la médiane sont très proches.

En revanche, si elle est dissymétrique (c'est le cas en présence de valeurs atypiques), la médiane sera inférieure (cas d'une distribution étalée à droite) ou supérieure (cas d'une distribution étalée à gauche) à la moyenne.

Remarque 2 : Les quantiles doivent être des valeurs de l'échantillon (le cas d'un échantillon discret)

Statistiquement, le quantile peut ou non, être une valeur de l'échantillon. Toutefois, par convention, il sera toujours une valeur de l'échantillon.

Considérons, à titre d'exemple, deux échantillons de valeurs et supposons qu'on souhaite obtenir la médiane de chacun d'eux :

- l'échantillon 1 $\{1,2,3,4,5,6,7,8,9,10\}$
- l'échantillon 2 $\{1,2,3,4,5,6,7,8,9,10,11\}$.

- Le premier échantillon compte dix observations. La médiane est ici entre les observations 5 et 6, c'est-à-dire à la 5,5^{ème} observation dans l'ensemble des données. Elle n'est pas, ici, une valeur de l'échantillon
- Le second échantillon comporte onze observations. Il existe donc une solution unique, médiane de l'échantillon : $\{6\}$.

Toutefois, la médiane doit être une valeur de l'échantillon. Ainsi, elle sera définie :

- Soit statistiquement : Dans le cas de l'échantillon 2, il existe bien une valeur de l'échantillon qui est la valeur médiane : $\{6\}$.
- Soit par convention : Dans le cas de l'échantillon 1, on choisira le plus petit élément de l'intervalle, pour se conformer à la propriété de continuité à gauche de la fonction quantile. La valeur médiane est donc $\{5\}$.

Cette méthode peut être généralisée à l'ensemble des quantiles d'un échantillon discret (cf. (3)) :

- Soit il existe une seule observation de l'échantillon qui représente le quantile solution : $\hat{Q}_Y(\theta) = a(\theta)$ (c'est le cas si $n\theta$ n'est pas un nombre entier).
- Soit il y a un intervalle de solutions. Il existe donc un ensemble de solutions $\{y_i = a(\theta)\}$ (ceci est le cas si $n\theta$ est un nombre entier). La définition de la fonction quantile est telle qu'on choisira ici le plus petit élément de cet intervalle.

Remarque 3 : Le cas d'un échantillon continu

Dans le cas d'un échantillon continu, il y a, dans la grande majorité des cas, une solution unique. Il existe une seule observation y_i telle que $y_i = a(\theta)$ et telle que $n\theta$ observations se situent sous ce quantile. $\hat{Q}_Y(\theta) = a(\theta)$ est la seule solution possible.

La détermination de la médiane ou de tout autre quantile ne pose alors aucun problème : en effet, toute valeur de la fonction de répartition F admet un antécédent unique.

4. Le point de départ des régressions quantiles : les quantiles non conditionnels sont solution d'un programme d'optimisation

Les quantiles *non conditionnels* d'un échantillon s'obtiennent par simple ordonnancement des observations.

Koenker et Bassett (1978) ont cependant montré que ces quantiles pouvaient aussi être obtenus à partir d'un programme d'optimisation, généralisable, par la suite, à l'estimation des quantiles *conditionnels* et plus précisément, des quantiles conditionnels ayant une forme linéaire.

Il s'agit tout d'abord de décrire le programme d'optimisation permettant d'estimer les différents quantiles non conditionnels, puis de montrer comment il peut être résolu. La prochaine section généralisera cette méthode à l'estimation des quantiles conditionnels.

A. Écriture du programme d'optimisation

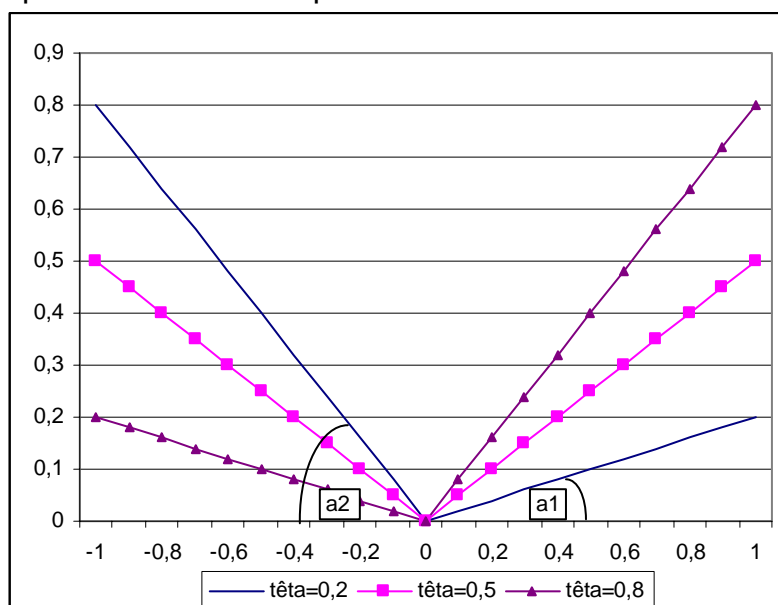
Introduction de la « fonction de perte »

Koenker et Bassett (1978) introduisent la « check function » ou « quantile loss function » que nous traduirons par « fonction de perte ».

Elle est définie par $\rho_{\theta}(y) = y(\theta - I(y \leq 0))$.

C'est une fonction convexe (car c'est une somme de fonctions convexes) et linéaire par morceaux. Elle n'est pas dérivable en l'origine.

Graphique 3 : Trois fonctions de perte associées à des valeurs de θ différentes



Trois fonctions de perte associées à trois valeurs différentes de θ sont ici représentées.

Pour $y \geq 0$, $\rho_\theta(y) = \theta y$ et la pente de la fonction de perte est $Tg(a_1) = \theta$.

Pour $y \leq 0$, $\rho_\theta(y) = (\theta - 1)y$ et la pente de la fonction de perte est $Tg(a_2) = \theta - 1$.

Cette fonction de perte a un rôle important dans la détermination des quantiles (non conditionnels) puisqu'elle permet de simplifier l'écriture de la fonction objectif.

La fonction objectif

L'estimation de la médiane non conditionnelle :

La médiane non conditionnelle est solution du programme de minimisation suivant :

$$\boxed{\text{Min}_{q_m \in \mathbb{R}} \sum_{i=1}^n |y_i - q_m|} \quad (6)$$

La médiane est donc la valeur y_i de l'échantillon qui minimise une somme d'écart absolu aux observations.

Démonstration

$$\text{On transforme (6) : } \sum_{i=1}^n |y_i - q_m| = \sum_{y_i \geq q_m} (y_i - q_m) + \sum_{y_i < q_m} (q_m - y_i)$$

Soit N_1 le nombre d'observations telles que $y_i \geq q_m$ et N_2 le nombre d'observations telles que $y_i < q_m$.

Alors :

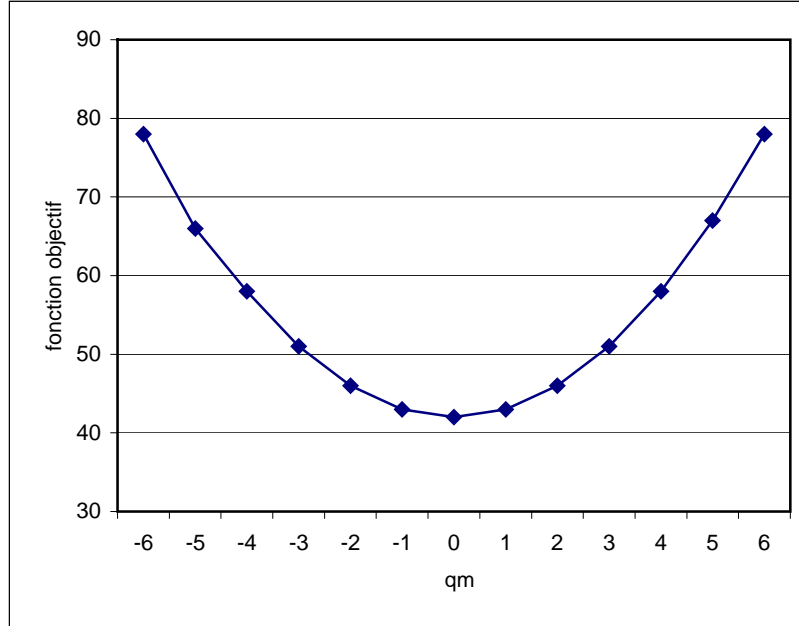
$$\begin{aligned} \sum_{y_i \geq q_m} (y_i - q_m) + \sum_{y_i < q_m} (q_m - y_i) &= \sum_{y_i \geq q_m} y_i - N_1 q_m + N_2 q_m - \sum_{y_i < q_m} y_i \\ &= q_m (N_2 - N_1) + \left(\sum_{y_i \geq q_m} y_i - \sum_{y_i < q_m} y_i \right) \end{aligned}$$

Étant donnée cette re-formulation, un exemple permet de comprendre la forme de la fonction objectif et la valeur de solution ⁴²:

Soit un échantillon prenant les valeurs $\{-6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\}$. La fonction objectif a la forme suivante :

⁴² Voir section suivante pour une description plus détaillée de la forme de la fonction objectif et des étapes de la minimisation.

Graphique 4 : Valeur de la fonction objectif en fonction de q_m



La fonction objectif est une fonction continue de q_m , constituée d'une succession de segments de droites. C'est une fonction décroissante, puis croissante (la pente est d'abord négative puis positive), elle passe donc par un minimum. Celui-ci est atteint lorsque la pente de la fonction objectif s'annule, pour la valeur de q_m telle que $N_1 = N_2$, c'est-à-dire lorsque q_m est la médiane de l'échantillon : $q_m = \hat{Q}(0,5)$. Dans l'échantillon, la médiane est $\{0\}$.

Remarque

Comme précédemment, lorsque le minimum est atteint sur tout un intervalle sur lequel la pente de la fonction objectif s'annule, il y a un intervalle de solutions.

La médiane sera la plus petite valeur de cet intervalle.

Généralisation à l'estimation de l'ensemble des quantiles non conditionnels :

Puisque la symétrie de la fonction valeur absolue conduit à l'estimation de la médiane, minimiser une somme *asymétriquement* pondérée de résidus absolus permet d'estimer les autres quantiles d'un échantillon.

Le $\theta^{\text{ème}}$ quantile non conditionnel de l'échantillon, noté $\hat{Q}_y(\theta)$, est solution du programme d'optimisation suivant :

$$\begin{aligned} \hat{Q}_y(\theta) &= \text{Min}_{a(\theta) \in \mathbb{R}} \left(\sum_{i|y_i \geq a(\theta)} \theta |y_i - a(\theta)| + \sum_{i|y_i < a(\theta)} (1-\theta) |y_i - a(\theta)| \right) \\ &= \text{Min}_{a(\theta) \in \mathbb{R}} \left(\sum_{i|y_i \geq a(\theta)} \theta (y_i - a(\theta)) + \sum_{i|y_i < a(\theta)} (\theta-1)(y_i - a(\theta)) \right) \end{aligned} \quad (7)$$

Le passage à la fonction de perte est un moyen de simplifier les expressions précédentes :

$$\hat{Q}_Y(\theta) = \underset{a(\theta) \in \mathbb{R}}{\text{Min}} \sum_{i=1}^n \rho_{\theta}(y_i - a(\theta)) \quad (8)$$

La fonction de perte mesure ici la perte encourue à l'utilisation de $a(\theta)$ plutôt que de y_i .

Il s'agit ici d'une somme asymétriquement pondérée : les observations y_i supérieures à la valeur prédite $a(\theta)$ (c'est-à-dire les « résidus » positifs, le premier terme de la somme) sont pondérées par le poids θ du quantile à obtenir, alors que les observations y_i inférieures à $a(\theta)$ (c'est-à-dire les « résidus » négatifs, deuxième terme de la somme) sont pondérées par $(\theta - 1)$ ⁴³. Remarquons que lorsque $\theta = 0,5$ (c'est-à-dire pour estimer la médiane), un poids identique est accordé aux écarts positifs et négatifs.

La pondération utilisée est inversement proportionnelle au « poids » que représentent les deux termes dans la somme totale. En effet, pour l'estimation du $\theta^{\text{ème}}$ quantile de l'échantillon, il y aura $(1 - \theta)\%$ de résidus positifs et $\theta\%$ de résidus négatifs. Le critère de minimisation est donc satisfait quand $\theta\%$ des résidus sont négatifs.

Éléments d'interprétation

La fonction à minimiser pour aboutir à l'estimation du $\theta^{\text{ème}}$ quantile de l'échantillon est :

$$\hat{Q}_Y(\theta) = \underset{a(\theta) \in \mathbb{R}}{\text{Min}} \left(\underbrace{\sum_{i|y_i \geq a(\theta)} \theta(y_i - a(\theta))}_{\text{"somme de gauche"}} + \underbrace{\sum_{i|y_i < a(\theta)} (\theta - 1)(y_i - a(\theta))}_{\text{"somme de droite"}} \right) \quad (9)$$

Simulons un échantillon de vingt observations ordonnées (y_1, \dots, y_{20}) , avec $y_i \rightarrow U(0,10)$. Le programme de minimisation précédent consiste à trouver, parmi les vingt observations candidates, l'observation y_i telle que la fonction est minimale. On cherche par exemple ici à trouver le 3^{ème} quartile de l'échantillon ($\theta=0,75$).

La valeur des observations, ainsi que les valeurs de la fonction objectif calculées lorsque chaque y_i est considéré comme solution du programme de minimisation, sont retranscrites dans le tableau ci-dessous :

⁴³ Par exemple, si $\theta = 0,2$, 80% du poids total est accordé aux résidus négatifs. En revanche, si $\theta = 0,8$, ce sont les résidus positifs qui reçoivent 80% du poids total

Tableau 1 : Les différentes valeurs prises par la fonction objectif lors des étapes de la minimisation

Observations	Valeur de y_i	Somme de gauche	Somme de droite	Fonction Objectif
1	-14,29	196,92	0	196,92
2	-12,31	168,73	0,49	169,23
3	-10,82	148,59	1,24	149,83
4	-10,00	138,07	1,86	139,93
5	-8,31	117,82	3,55	121,36
6	-7,24	105,81	4,88	110,69
7	-6,49	97,88	6,01	103,89
8	-3,32	67,01	11,55	78,56
9	-3,17	65,66	11,85	77,52
10	-2,74	62,12	12,82	74,94
11	-0,58	45,94	18,21	64,16
12	0,12	41,16	20,16	61,32
13	0,97	36,11	22,68	58,80
14	3,41	23,27	30,64	53,90
15	<u>5,49</u>	<u>13,94</u>	<u>37,89</u>	<u>51,83</u>
16	<u>6,42</u>	<u>10,42</u>	<u>41,41</u>	<u>51,83</u>
17	6,83	9,21	43,02	52,23
18	8,71	4,97	51,04	56,01
19	9,59	3,65	54,98	58,63
20	14,46	0	78,12	78,12

Étape 1 : Si on suppose que l'observation y_1 est solution, alors le programme précédent se réécrit :

$$\hat{Q}_Y(\theta) = \underset{a(\theta) \in \mathbb{R}}{\text{Min}} \left(\sum_{y_i \geq y_1} \theta(y_i - y_1) \right).$$

Les 20 observations ont toutes un poids θ .

Étape 2 : En revanche, si on suppose que c'est l'observation y_2 qui est solution, le programme devient :

$$\hat{Q}_Y(\theta) = \underset{a(\theta) \in \mathbb{R}}{\text{Min}} \left(\sum_{y_i \geq y_2} \theta(y_i - y_2) + (\theta - 1)(y_1 - y_2) \right)$$

Le poids de l'observation y_1 est passé de θ à $\theta - 1$, soit une perte de la contribution de y_1 de 1. Le poids des observations (y_2, \dots, y_{20}) est resté de θ .

Remarquons que le candidat potentiel ne « compte » pas dans la fonction objectif, puisque le terme de la somme qui lui correspond est nul.

Étape k : Si l'observation y_k est solution, le programme est alors :

$$\hat{Q}_Y(\theta) = \text{Min}_{\theta \in \mathbb{R}} \left(\sum_{y_i \geq y_k} \theta(y_i - y_k) + \sum_{y_i < y_k} (\theta - 1)(y_i - y_k) \right)$$

Le poids des observations $y_i < y_k$ est passé de θ à $\theta - 1$. Le poids de toutes les observations (y_k, \dots, y_{20}) est resté de θ .

Si y_i^* est la valeur de l'échantillon solution du programme, alors le poids accordé à tous les $y_i < y_i^*$ change de -1 (passe de θ à $\theta - 1$).

À chaque étape de la minimisation (c'est-à-dire au fur et à mesure que les différentes observations candidates y_i sont testées, par ordre croissant) :

- un terme positif ou nul (avec le poids θ) est supprimé du membre de gauche de la fonction objectif
- un terme négatif est rajouté à la somme de droite de la fonction objectif (avec le poids $\theta - 1$)

Le graphique suivant permet de mieux comprendre le comportement de la fonction objectif à chaque étape de la minimisation (graphique 5).

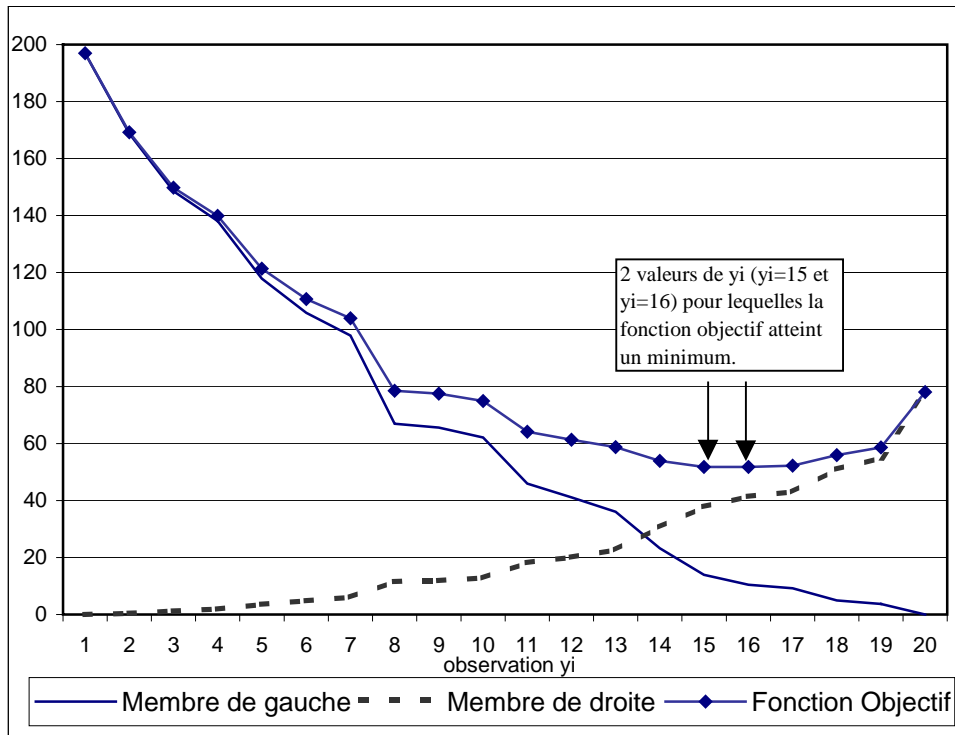
La courbe notée « gauche » est la représentation du membre de gauche de (9). Elle est décroissante car à chaque étape de la minimisation, on supprime un terme positif ou nul. Par ailleurs, la courbe « droite », représentative du membre de droite de cette même fonction objectif est croissante, car à chaque étape, c'est un terme positif qui est rajouté.

Le graphique montre que la fonction objectif est convexe (somme de fonctions convexes) et linéaire par morceaux.

Sa pente passe de $-\sum_i \theta$ (pour $y_i < y_1$) à $\sum_i (1 - \theta)$ (pour $y_i > y_{20}$).

Lorsque y_i augmente de y_1 à y_{20} , la pente de la fonction passe de négative à positive. Elle passe donc bien par un minimum. L'optimalité se situe au minimum de la fonction objectif : celle-ci est croissante à gauche et à droite de $\hat{Q}_y(\theta)$.

Graphique 5 : Graphique de la fonction objectif et modification de la somme « de gauche » et « de droite » au cours de la minimisation



B. Résolution analytique du programme de minimisation dans le cas des quantiles non conditionnels théoriques

Il s'agit maintenant de démontrer analytiquement que l'observation en laquelle la fonction objectif atteint un minimum est bien $\hat{Q}_y(\theta)$.

Soit une variable aléatoire Y .

Conformément aux programmes décrits jusqu'à présent, le $\theta^{\text{ème}}$ quantile de Y est donc solution du programme d'optimisation

$$\boxed{\text{Min}_{a(\theta)} E[\rho_\theta(Y - a(\theta))]} \quad (10)$$

Cette fonction de perte espérée doit être minimisée par rapport à $a(\theta)$: on cherche donc le $a(\theta)$ pour laquelle cette fonction atteint un minimum. C'est le cas lorsque $a(\theta) = \hat{Q}_y(\theta)$, le $\theta^{\text{ème}}$ quantile de Y .

Démonstration

$$\begin{aligned} E[\rho_\theta(Y - a(\theta))] &= E[\theta(Y - a(\theta)) + (\theta - 1)(Y - a(\theta))] \\ &= \int_{a(\theta)}^{\infty} \theta(y - a(\theta))dF(y) + \int_{-\infty}^{a(\theta)} (\theta - 1)(y - a(\theta))dF(y) \end{aligned}$$

Alors :

$$\frac{\partial E(\rho_\theta(Y - a(\theta)))}{\partial a(\theta)} = \theta \frac{\partial \int_{a(\theta)}^{+\infty} (y - a(\theta))dF(y)}{\partial a(\theta)} + (\theta - 1) \frac{\partial \int_{-\infty}^{a(\theta)} (y - a(\theta))dF(y)}{\partial a(\theta)} \quad (11)$$

Les bornes de l'intervalle dépendant de la variable par rapport à laquelle se fait la dérivation, la règle de Leibniz doit être utilisée.

Règle de Leibniz

Soit une fonction $Q(x, y)$ et des bornes $\alpha(x)$ et $\beta(x)$. La règle d'intégration de Leibniz est la suivante :

$$\int_{\alpha(x)}^{\beta(x)} \frac{\partial}{\partial x} Q(x, y) dy = \frac{\partial}{\partial x} \int_{\alpha(x)}^{\beta(x)} Q(x, y) dy - Q(x, \beta(x)) \frac{\partial \beta(x)}{\partial x} - Q(x, \alpha(x)) \frac{\partial \alpha(x)}{\partial x} \quad (12)$$

Ici, ce sont des quantiles non conditionnels qui sont estimés donc les termes correctifs disparaissent car les bornes sont des constantes. Simplifions chaque terme de la somme définie en (11) grâce à la règle (12) :

$$\begin{aligned} \theta \frac{\partial \int_{a(\theta)}^{+\infty} (y - a(\theta))dF(y)}{\partial a(\theta)} &= \theta \left[\int_{a(\theta)}^{+\infty} \frac{\partial (y - a(\theta))dF(y)}{\partial a(\theta)} - (a(\theta) - a(\theta)) \frac{\partial a(\theta)}{\partial a(\theta)} \right] \\ &= \theta \int_{a(\theta)}^{+\infty} \frac{\partial (y - a(\theta))dF(y)}{\partial a(\theta)} \\ &= -\theta \int_{a(\theta)}^{+\infty} dF(y) \\ &= -\theta(1 - F(a(\theta))) \end{aligned}$$

Et :

$$\begin{aligned}
 (\theta - 1) \frac{\partial \int_{-\infty}^{a(\theta)} (y - a(\theta)) dF(y)}{\partial a(\theta)} &= (\theta - 1) \left[\int_{-\infty}^{a(\theta)} \frac{\partial (y - a(\theta)) dF(y)}{\partial a(\theta)} + (a(\theta) - a(\theta)) \frac{\partial a(\theta)}{\partial a(\theta)} \right] \\
 &= (\theta - 1) \int_{-\infty}^{a(\theta)} \frac{\partial (y - a(\theta)) dF(y)}{\partial a(\theta)} \\
 &= -(\theta - 1) \int_{-\infty}^{a(\theta)} dF(y) \\
 &= -(\theta - 1) F(a(\theta))
 \end{aligned}$$

Donc

$$\begin{aligned}
 \frac{\partial E(\rho_\theta(Y - a(\theta)))}{\partial a(\theta)} &= -\theta [1 - F(a(\theta))] - (\theta - 1) F(a(\theta)) \\
 &= -\theta + \theta F(a(\theta)) - \theta F(a(\theta)) + F(a(\theta)) \\
 &= -\theta + F(a(\theta))
 \end{aligned}$$

Cette fonction atteint un minimum si :

$$\boxed{-\theta + F(a(\theta)) = 0 \Leftrightarrow \theta = F(a(\theta))}$$

La fonction objectif est donc minimisée si $a(\theta)$ est tel que $F(a(\theta)) = \theta$, c'est-à-dire si la solution $a(\theta)$ est le $\theta^{\text{ème}}$ quantile de la distribution non conditionnelle de Y , défini en (2).

Par ailleurs, puisque la fonction de perte est convexe (car c'est une somme de fonctions convexes), sa dérivée seconde est positive. Les conditions du second ordre sont donc vérifiées et la solution obtenue est bien un minimum.

C. Conséquences et intuitions

- Il est possible de généraliser les notations précédentes à l'estimation des quantiles non conditionnels empiriques, basés sur un échantillon (y_1, \dots, y_n) . Ils sont solutions du programme :

$$\boxed{\hat{Q}_Y(\theta) = \underset{a(\theta) \in \mathbb{R}}{\text{Min}} \sum_{i=1}^n \rho_\theta(y_i - a(\theta))} \quad (13)$$

- Par ailleurs, pour estimer des quantiles conditionnels, le même type de programme peut être posé.

Nous entendons ici par quantile conditionnel l'estimation de quantiles conditionnels linéaires.

Ainsi, il s'agit d'estimer la fonction quantile conditionnelle linéaire $Q_Y(\theta|x) = x_i' \beta(\theta)$ en résolvant le programme : $\hat{\beta}(\theta) = \text{Min}_{\beta} \sum_{i=1}^n |y_i - x_i' \beta(\theta)|$.

Par exemple, l'estimateur des moindres écarts absolus (MEA) est solution du programme $\hat{\beta}(0,5) = \text{Min}_{\beta} \sum_{i=1}^n |y_i - x_i' \beta(\theta)|$. Cet estimateur est aussi connu sous le nom d'estimateur L1, d'estimateur de la régression médiane ou, en anglais, de « LAD estimator » (least absolute deviations estimator).

Il faut ici noter l'analogie avec l'estimateur des moindres carrés ordinaires (MCO), solution du programme $\text{Min}_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2$.

Avec les MCO, il s'agit d'ajuster une droite au nuage de points, en minimisant la somme des résidus au carré. La régression médiane consiste également à ajuster une droite au nuage de points, mais cet ajustement se fait en minimisant la somme des résidus absolus.

Remarque

Si la distribution est symétrique, l'estimateur des Moindres Carrés Ordinaires et des Moindres Écarts Absolus coïncident car la moyenne et la médiane sont identiques.

L'objectif de Koenker et Bassett est donc de généraliser les programmes précédents au cadre des régressions quantiles. En effet, la formulation des quantiles non conditionnels comme solution d'un programme d'optimisation permet d'étendre cette méthode à l'estimation des quantiles conditionnels, et en particulier de généraliser les quantiles dans un contexte de régression.

Il s'agira alors de résoudre le programme de minimisation :

$$\boxed{\hat{\beta}(\theta) = \text{Min}_{\beta} \sum_{i=1}^n \rho_{\theta}(y_i - x_i' \beta(\theta))} \quad (14)$$

La section 5 décrit plus précisément les hypothèses du modèle et la mise en œuvre des régressions quantiles.

5. Généralisation à l'estimation des régressions quantiles

A. Hypothèses du modèle

Soit $(y_i, x_i), i = 1 \dots n$ un échantillon, avec x_i un vecteur $P \times 1$ de variables explicatives (contenant aussi la constante) et y_i la variable expliquée.

$\beta = (\beta_1, \dots, \beta_p)'$ est le vecteur $P \times 1$ des paramètres à estimer et $u = (u_1, \dots, u_n)'$ le vecteur $N \times 1$ des perturbations.

Supposons que $y_i = x_i' \beta(\theta) + u_i(\theta)$ (15)

On suppose donc une relation linéaire la variable expliquée y et les variables explicatives x ⁴⁴.

À la différence des moindres carrés ordinaires, la distribution du terme d'erreur, $u_i(\theta) = y_i - x_i' \beta(\theta)$ est inconnue. On suppose seulement que le $\theta^{\text{ème}}$ quantile de la distribution des erreurs conditionnellement aux variables explicatives est nul :

$$Q_{u_i(\theta)}(\theta | x_i) = 0 \quad (16)$$

La fonction quantile conditionnelle de Y sachant x est notée :

$$Q_Y(\theta | x) = x_i' \beta(\theta) \quad (17)$$

Démonstration

Soit un échantillon de variables aléatoires (Y_1, \dots, Y_n) distribuées telles que $F(y|x) = \Pr(Y_i \leq y|x)$.

Notations

- $Q_Y(\theta|x)$ est le $\theta^{\text{ème}}$ quantile de Y sachant x .
- $F_{u_i(\theta)}(\cdot|x)$ est la fonction de répartition de $u_i(\theta)$ sachant x .

$$\begin{aligned} F(y|x) &= \Pr(Y_i \leq y|x_i) = \Pr(x_i' \beta(\theta) + u_i(\theta) \leq y|x_i) \\ &= \Pr(u_i(\theta) \leq y - x_i' \beta(\theta)|x_i) \\ &= F_{u_i(\theta)}(y - x_i' \beta(\theta)|x_i), i = 1 \dots n \end{aligned}$$

⁴⁴ Cette hypothèse est restrictive mais elle peut facilement être relâchée par l'introduction de variables muettes, d'une fonction polynomiale des variables explicatives ou d'interaction entre les variables explicatives.

Par ailleurs :

$$\begin{aligned}
Q_Y(\theta|x_i) &= \inf \{y \text{ t.q. } F_Y(y|x) \geq \theta\} \\
&= \inf \{y \text{ t.q. } F_{u_i(\theta)}(y - x_i' \beta(\theta)|x_i) \geq \theta\} \\
&= \inf \{y \text{ t.q. } (y - x_i' \beta(\theta)) \geq F_{u_i(\theta)}^{-1}(\theta|x_i)\} \\
&= \inf \{y \text{ t.q. } y \geq x_i' \beta(\theta) + F_{u_i(\theta)}^{-1}(\theta|x_i)\} \\
Q_Y(\theta|x_i) &= x_i' \beta(\theta) + F_{u_i(\theta)}^{-1}(\theta|x_i)
\end{aligned}$$

Si on fait l'hypothèse que $F_{u_i(\theta)}^{-1}(\theta|x_i) = Q_{u_i(\theta)}(\theta|x_i) = 0$, alors le θ^{ime} quantile conditionnel de Y sachant x est défini comme en (17)⁴⁵.

B. Programme d'optimisation

Pour tout $\theta \in [0,1]$, $\hat{\beta}(\theta)$ est solution de :

$$\begin{aligned}
\hat{\beta}(\theta) &= \text{Min}_{\beta \in \mathbb{R}^p} \left(\sum_{y_i | y_i \geq x_i' \hat{\beta}(\theta)} \theta |y_i - x_i' \beta(\theta)| + \sum_{y_i | y_i < x_i' \hat{\beta}(\theta)} (1-\theta) |y_i - x_i' \beta(\theta)| \right) \\
&= \text{Min}_{\beta \in \mathbb{R}^p} \left(\sum_{y_i | y_i \geq x_i' \hat{\beta}(\theta)} \theta (y_i - x_i' \beta(\theta)) + \sum_{y_i | y_i < x_i' \hat{\beta}(\theta)} (1-\theta) (y_i - x_i' \beta(\theta)) \right) \quad (18) \\
&= \text{Min}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\theta(y_i - x_i' \beta(\theta))
\end{aligned}$$

Comme précédemment, les $(1-\theta)$ observations telles que $y_i \geq x_i' \beta(\theta)$ (les résidus positifs) sont pondérées par θ et les θ observations telles que $y_i < x_i' \beta(\theta)$ (les résidus négatifs) sont pondérées par $(\theta-1)$.

L'estimateur des régressions quantiles n'est pas facile à calculer, il n'y a pas de solution explicite comme dans le cas de moindres carrés ordinaires.

Cela a constitué le principal frein au développement de la méthode. En effet, la fonction objectif, linéaire par morceaux et continue, n'est pas différentiable aux quelques points auxquels le ou les résidus sont nuls.

Les méthodes d'optimisation numériques standard (basées sur les dérivées des fonctions objectif) ne sont donc pas applicables immédiatement.

⁴⁵ Il faut noter ici la ressemblance avec l'estimation par moindres carrés ordinaires. Soit $y_i = x_i' \beta + u_i$ un modèle de régression linéaire. Si on fait l'hypothèse que $E(u_i|x_i) = 0$, alors l'espérance conditionnelle de Y sachant x est $E(y_i|x_i) = x_i' \beta$.

Les régressions quantiles sont donc mises en œuvre par des méthodes de programmation linéaire : le programme de minimisation précédent doit donc être reformulé comme un programme linéaire⁴⁶ (cf. Buchinsky (1998) par exemple).

La programmation linéaire

De manière générale, les programmes à résoudre sont du type :

$\begin{array}{l} \text{Min } Cx \\ \text{s.c. } Ax = b \\ x \geq 0 \end{array}$
--

où x est inconnue, A est une matrice de coefficients certaine, c et b sont des vecteurs de coefficients connus. La forme standard de programmation linéaire demande à toutes les variables d'être positives ou nulles.

L'algorithme du simplexe est la méthode la plus couramment utilisée pour résoudre ce type de programmes.

Résolution

L'équation à estimer est notée $y_i = x_i' \beta(\theta) + u_i(\theta) = \sum_{j=1}^P x_{ij} \beta_j(\theta) + u_i(\theta)$

Il est possible de réécrire y_i comme une fonction d'éléments positifs (en introduisant des « variables artificielles »), puis de transformer l'écriture du modèle ci-dessus en écriture matricielle.

Les variables artificielles introduites sont les suivantes :

- $\beta^+(\theta) = \max(\beta_j, 0) \geq 0$ représente la partie positive du vecteur de coefficients β_j
- $\beta^-(\theta) = -\min(\beta_j, 0) \geq 0$ représente la partie négative du vecteur de coefficients β_j
 - On a $\beta_j = \beta_j^+ - \beta_j^-$
- $u_i^+(\theta) = \max\left(\left(y_i - \sum_{j=1}^P x_{ij}' \beta_j(\theta)\right), 0\right) \geq 0$ représente la partie positive du vecteur des résidus
- $u_i^-(\theta) = -\min\left(\left(y_i - \sum_{j=1}^P x_{ij}' \beta_j(\theta)\right), 0\right) \geq 0$ représente la partie négative du vecteur des résidus
 - On a $u_i = u_i^+ - u_i^-$

⁴⁶ En pratique, l'estimation des régressions quantiles grâce aux logiciels, STATA ou R se réalise très facilement, grâce à une commande simple (cf. section 11 pour une présentation des commandes sous STATA).

Le programme de minimisation initial peut être transformé :

$$P = \text{Min} \sum_{i=1}^n \rho_{\theta} \left(y_i - \sum_{j=1}^p x_{ij} \beta_j(\theta) \right)$$

$$= \text{Min} \sum_{i=1}^n \rho_{\theta} (u_i(\theta))$$

Or :

$$\rho_{\theta}(u_i(\theta)) = \theta u_i(\theta) \text{ si } u_i(\theta) \geq 0$$

$$= \theta u_i^+(\theta) \text{ si } u_i(\theta) \geq 0$$

et

$$\rho_{\theta}(u_i(\theta)) = (\theta - 1)u_i(\theta) \text{ si } u_i(\theta) < 0$$

$$= (1 - \theta)u_i^-(\theta) \text{ si } u_i(\theta) < 0$$

Donc

$$P = \text{Min} \sum_{i=1}^n (\theta u_i^+(\theta) + (1 - \theta)u_i^-(\theta))$$

Le programme de minimisation est donc :

$\text{Min} \sum_{i=1}^n (\theta u_i^+(\theta) + (1 - \theta)u_i^-(\theta))$ $\text{s.c. } y_i = \sum_{j=1}^p x_{ij} [\beta_j^+(\theta) - \beta_j^-(\theta)] + [u_i^+(\theta) - u_i^-(\theta)] \quad (19)$ $\beta_j^+(\theta) \geq 0, \beta_j^-(\theta) \geq 0, u_i^+(\theta) \geq 0, u_i^-(\theta) \geq 0$

Sous forme matricielle, ce programme peut se réécrire :

$\text{Min } C' z$ $\text{s.c. } Az = y$ $\text{et } z \geq 0$
--

avec :

- $C = (0', 0', \theta l', (1 - \theta)l')$ où 0 est un vecteur de p lignes de 0 et l un vecteur de n lignes de 1 .
- $z = (\beta^+(\theta)', \beta^-(\theta)', u_i^+(\theta)', u_i^-(\theta)')$ où $\beta^+(\theta)'$ et $\beta^-(\theta)'$ sont deux vecteurs de dimension $(p,1)$ et $u_i^+(\theta)'$ et $u_i^-(\theta)'$ deux vecteurs de dimension $(n,1)$.
- $A_{(n, 2p+2n)} = (X, -X, I_n, -I_n)$ où X est la matrice de dimension (n,p) des variables explicatives et I_n la matrice identité de format (n,n) .
- $y_{(n,1)} = (y_1, \dots, y_n)'$ est la matrice de la variable à expliquer.

La résolution de ce programme conduit à la solution : $X\beta(\theta) = F^{-1}(\theta) = Q_Y(\theta|X)$

Ou encore, si les écritures matricielles sont abandonnées : $Q_Y(\theta|x_i) = x_i'\beta(\theta)$

C. Conséquences

Les régressions quantiles

Chaque régression quantile caractérise quelques points de la distribution conditionnelle. Ainsi, l'estimation de cinq régressions quantiles, avec $\theta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ ne permet pas de caractériser l'ensemble de la distribution conditionnelle de Y sachant $X = x$.

C'est la combinaison d'un grand nombre de régressions quantiles, avec $\theta \in [0,1]$ qui permet de décrire entièrement la distribution conditionnelle sous-jacente.

Remarque

Si un échantillon d'observations est partagé en dix sous-parties, selon les déciles non conditionnels de la variable expliquée, appliquer les moindres carrés ordinaires sur chacun de ces sous-groupes ne revient pas à l'estimation de dix régressions quantiles (pour $\theta = \{0.1, 0.2, \dots, 0.9\}$). En effet, de tels moindres carrés ordinaires n'utilisent, pour l'estimation de y , qu'une partie des observations de l'échantillon, conduisant à un biais de sélection (Heckman, 1979). Par ailleurs, si les déciles extrêmes sont en général assez bien estimés, en raison de la forte variabilité des observations dans ces déciles, ce ne sera pas le cas pour l'ensemble des déciles intermédiaires.

Interprétation des coefficients

Les régressions quantiles estiment l'influence des variables explicatives à différents points de la distribution conditionnelle de Y .

Par exemple, dans la partie précédente, c'est l'influence des variables sexe, ancienneté et densité médicale aux différents quantiles de la distribution conditionnelle du logarithme des honoraires qui a été étudiée.

Interpréter les coefficients estimés des différentes régressions quantiles revient à commenter l'influence d'une modification d'une des variables explicatives sur le $\theta^{\text{ème}}$ quantile conditionnel du logarithme des honoraires, toutes choses égales par ailleurs.

Dans l'étude précédente, cela ne signifie cependant pas qu'un omnipraticien se situant au $\theta^{\text{ème}}$ quantile de la distribution conditionnelle des honoraires appartiendra toujours à ce même quantile après la variation d'une de ses caractéristiques personnelles.

Par ailleurs, le raisonnement porte sur la distribution *conditionnelle* (des honoraires dans notre cas).

De ce fait, le bas de la distribution conditionnelle des honoraires (par exemple la dixième ou vingtième régression quantile) ne se réfère pas obligatoirement aux omnipraticiens ayant de faibles honoraires. En revanche, il se réfère aux omnipraticiens qui, conditionnellement à leurs caractéristiques personnelles (sexe, ancienneté, densité d'omnipraticiens libéraux dans le département d'exercice), ont des honoraires plus faibles que ce qu'on pourrait attendre, c'est-à-dire en comparaison de leur groupe socio-démographique. On peut dire que ce sont les omnipraticiens les moins « productifs ».

En revanche, ceux du haut de la distribution (par exemple l'estimation de la quatre-vingtième ou quatre-vingt-dixième régression quantile) peuvent être considérés comme les plus « productifs » : leurs caractéristiques personnelles inobservables (goût pour le travail dans l'arbitrage travail/loisir, capacité à fidéliser la clientèle, qualités relationnelles) leur permettent d'avoir des honoraires plus élevés que ce que prédisent leurs caractéristiques observables.

Si plusieurs régressions quantiles sont estimées, le raisonnement *conditionnellement à X* permet de regarder, pour une même valeur $X = x$, les écarts dans les différents quantiles estimés de Y .

Ces écarts mesurent donc les différences dans les valeurs de Y qui proviennent de différences dans les caractéristiques *inobservables* (ou non prises en compte dans les X) entre les individus.

Retour sur les propriétés des perturbations

La solution du programme d'optimisation est telle que $Q_{u_i(\theta)}(\theta|x_i) = 0$.

Il apparaît que si le modèle contient p variables explicatives (dont la constante), il y a exactement p observations telles que $Q_{u_i(\theta)}(\theta|x_i) = 0$. La $\theta^{\text{ème}}$ régression quantile passe exactement par p observations de l'échantillon.

Par ailleurs, il y aura approximativement $n\theta$ résidus négatifs, c'est-à-dire environ $n\theta$ observations sous le $\theta^{\text{ème}}$ quantile estimé et approximativement $n(1-\theta)$ résidus positifs, c'est-à-dire $n(1-\theta)$ observations au-dessus du $\theta^{\text{ème}}$ quantile estimé.

L'estimation de la 25^{ème} régression quantile est donc telle qu'il y aura 25 % de résidus négatifs et 75 % de résidus positifs.

Le pseudo-R² : une mesure de la qualité de l'ajustement

Dans l'estimation par moindres carrés ordinaires, le R² (coefficient de détermination) mesure la dispersion des observations autour de l'hyperplan qui représente le modèle estimé : plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement global du modèle.

Il est défini ainsi :

$$R^2 = 1 - \frac{\sum_i (y_i - x_i \hat{\beta})^2}{\sum_i (y_i - \bar{y})^2} \in [0,1]$$

Il peut être obtenu à partir de l'équation d'analyse de la variance, à savoir :

$$SCT = SCE + SCR$$

$$\sum_i (y_i - \bar{y})^2 = \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i u_i^2$$

La variation totale à expliquer (somme des carrés totale, SCT) peut se décomposer en somme des carrés expliquée (SCE) par le modèle et en somme des carrés des résidus (SCR), partie que le modèle n'explique pas.

Alors

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

Il est également possible de définir un critère d'ajustement du modèle avec les régressions quantiles. Il est noté « pseudo- R^2 », encore noté $R^2(\theta)$ et défini ainsi (Koenker et Machado, 1999) :

$$pseudo - R^2 = R^2(\theta) = 1 - \frac{\text{Min} \sum_i \rho_\theta(y_i - x_i \beta(\theta))}{\text{Min} \sum_i \rho_\theta(y_i - Q_Y(\theta))} \in [0,1] \quad (20)$$

Avec $Q_Y(\theta)$ le $\theta^{\text{ème}}$ quantile non conditionnel de Y .

1. $\hat{\beta}(\theta)$ est l'estimateur du $\theta^{\text{ème}}$ quantile de régression. Il est solution de

$$\hat{V}(\theta) = \text{Min}_b \sum_{i=1}^n \rho_\theta(y_i - x_i \beta)$$

2. $Q_Y(\theta)$ est le $\theta^{\text{ème}}$ quantile non conditionnel de Y . Il est solution de :

$$\tilde{V}(\theta) = \text{Min}_b \sum_{i=1}^n \rho_\theta(y_i - Q_Y(\theta))$$

3. On peut enfin définir la somme pondérée estimée des écarts absolus :

$$\bar{V}(\theta) = \text{Min}_b \sum_{i=1}^n \rho_\theta(x_i \hat{\beta}(\theta) - Q_Y(\theta))$$

Alors

$$R^2(\theta) = \frac{\bar{V}(\theta)}{\tilde{V}(\theta)} = 1 - \frac{\hat{V}(\theta)}{\tilde{V}(\theta)}$$

Le pseudo- R^2 n'est pas directement comparable au R^2 obtenu avec les moindres carrés ordinaires.

À la différence du R^2 , le pseudo R^2 n'est pas un critère d'ajustement global du modèle, mais d'ajustement local pour le $\theta^{ème}$ quantile, c'est-à-dire pour un quantile particulier, parmi tous les quantiles estimés.

Les droites de régressions estimées peuvent-elles se croiser ?

Pour $\theta_1 \leq \theta_2$, a t-on $\hat{Q}_Y(\theta_1|x) \leq \hat{Q}_Y(\theta_2|x), \forall x$?

$\hat{Q}_Y(\theta|x)$ est monotone en θ au voisinage du point moyen de l'échantillon, $\bar{x} = \frac{1}{n} \sum_i x_i$

(Koenker et Bassett (1982)). En dehors, il se peut que les fonctions quantiles estimées se croisent : $\hat{Q}_Y(\theta|x)$ n'est donc pas une fonction monotone de θ pour toutes les valeurs de x .

Pour une valeur $x = x_0$, le premier quartile conditionnel de la distribution des honoraires peut être supérieur au troisième quartile conditionnel. Cette situation n'est cependant pas fréquente et se produit surtout pour des valeurs de x pour lesquelles les résultats doivent être interprétés avec prudence (valeurs pour lesquelles il y a peu d'observations autour du quantile à estimer).

Dans l'estimation des honoraires en fonction de l'ancienneté de l'omnipraticien, ce cas se produit par exemple pour des valeurs de l'ancienneté supérieures à quarante ans, région dans laquelle tous les résultats obtenus doivent, de toutes façons, être interprétés avec précaution en raison du faible nombre d'omnipraticiens appartenant à cette tranche.

Quelles limites à la mise en œuvre des régressions quantiles ?

Ce sont les difficultés liées à sa mise en œuvre qui ont constitué le principal frein au développement des régressions quantiles.

En effet, dans le cas d'une estimation par moindres carrés ordinaires, il s'agit de chercher la valeur du paramètre β qui minimise : $Min_{\beta} \sum_{i=1}^n (y_i - x_i' \beta)^2$.

Cette fonction est dérivable.

En revanche, dans les cas des régressions quantiles, il s'agit de maximiser des fonctions non dérivables, comme nous l'avons montré dans la section 5. En pratique toutefois, les logiciels économétriques rendent la tâche plus facile.

6. Propriétés asymptotiques

A. Propriétés asymptotiques des quantiles de régressions

Soit le modèle $y_i = x_i' \beta + u_i$, avec u_i iid.

On note (Koenker et Bassett, 1978) F la fonction de répartition des u_i , f sa densité et on suppose que $f(F^{-1}(\theta)) > 0$. Par ailleurs, $R = \frac{1}{n} \sum_{i=1}^n x_i x_i'$ est une matrice semi-définie positive.

Soit $\hat{\beta}(\theta)$ l'estimateur de la $\theta^{\text{ème}}$ régression quantile. Sa distribution asymptotique est donnée par (Koenker et Bassett, théorème 4.2) :

$$\boxed{\sqrt{n}(\hat{\beta}(\theta) - \beta(\theta)) \rightarrow N\left(0, \frac{\theta(1-\theta)}{f^2(F^{-1}(\theta))} R^{-1}\right)} \quad (21)$$

Lorsque les erreurs ne sont pas iid, cette formule est plus complexe, et notamment car les densités f_i ne sont plus identiques. Alors :

$$\sqrt{n}(\hat{\beta}(\theta) - \beta(\theta)) \rightarrow N\left(0, \theta(1-\theta)H^{-1}JH^{-1}\right)$$

où $J(\theta) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n x_i x_i'$ et $H(\theta) = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n x_i x_i' f_i(\beta(\theta))$

$f_i(\beta(\theta))$ caractérise la densité conditionnelle de y_i évaluée au $\theta^{\text{ème}}$ quantile conditionnel.

Remarque 1

Le numérateur de la matrice de variance-covariance asymptotique fait penser à la variance d'une loi de Bernoulli : Si X est une variable aléatoire discrète (prenant les valeurs 0 et 1) suivant une loi de Bernoulli de paramètre θ , alors $V(X) = \theta(1-\theta)$.

Remarque 2

Dans le cadre des moindres carrés ordinaires, la matrice de variance covariance de l'estimateur $\hat{\beta}_{mco}$ est donnée par :

$$\boxed{V(\hat{\beta}_{mco}) = \sigma^2 (X'X)^{-1}} \quad (22)$$

Dès lors, deux éléments influencent la précision de l'estimation :

- La variance des perturbations, notée $\sigma^2 = \text{Var}(u_i)$, intervient au numérateur de la variance de l'estimateur. Celui-ci sera donc d'autant plus efficace que les perturbations sont peu dispersées.

- La matrice des variables explicatives : $(X'X)^{-1}$. Puisqu'elle intervient au dénominateur, plus les observations sont dispersées, plus les estimateurs seront précis.

Dans le cas des régressions quantiles, on observe également un arbitrage entre numérateur et dénominateur (cf. formule (21)).

- $\frac{\theta(1-\theta)}{f^2(F^{-1}(\theta))}$ prend le rôle de $\sigma^2 = \text{Var}(u_i)$: l'estimation des queues de distribution, c'est-à-dire θ proche de 0 ou 1 rend les estimations plus précises, car cette « variance » est plus faible (dans ce cas, le ratio est en effet proche de 0).
- $R-1$ prend le rôle de $(X'X)^{-1}$: plus les observations sont dispersées, plus les estimateurs seront précis.

Il apparaît que la précision asymptotique dépend de l'inverse de la fonction de densité, évalué au quantile considéré. Ce terme prend le nom de « sparsity function » (terme introduit par Tukey en 1965), ce qui peut être traduit par « fonction de rareté ».

L'estimation par moindres carrés ordinaires nécessite d'estimer la variance des perturbations qui est inconnue afin de connaître la matrice de variance-covariance de l'estimateur.

De façon similaire, c'est cette fonction de rareté qu'il faut estimer dans le cadre des régressions quantiles, afin d'estimer la matrice de variance-covariance asymptotique.

Cette procédure est difficile à mettre en œuvre, on préférera donc ici employer le bootstrap dont le principe est plus simple à comprendre (méthode d'autant plus rapide que les erreurs ne sont pas *iid*)⁴⁷.

B. Estimation de la matrice de variance-covariance asymptotique par Bootstrap :

Plusieurs algorithmes de Bootstrap peuvent être utilisés :

- le Bootstrap des résidus (suppose que les erreurs sont *iid*)
- le Bootstrap par paires (dans le cas où les erreurs ne sont pas *iid*)

Le Bootstrap des résidus

Soit un échantillon $(x_i, y_i), i = 1, \dots, n$. Cette méthode suppose que les erreurs sont *iid* et le rééchantillonnage est effectué sur le $\theta^{\text{ème}}$ résidus.

$\hat{\beta}(\theta)$ est estimé grâce à la $\theta^{\text{ème}}$ régression quantile et on en déduit $\hat{u}_i(\theta) = y_i - x_i \hat{\beta}(\theta)$. La fonction de répartition empirique estimée ($\hat{F}_n(u_i(\theta))$) permet de créer des résidus bootstrapés $u_i^*(\theta)$ et donc de générer des $y_i^* = x_i \hat{\beta}(\theta) + u_i^*(\theta)$.

⁴⁷ En pratique, lorsque les estimations sont réalisées avec le logiciel STATA, il n'y a qu'une option à rajouter à la procédure afin d'obtenir des écarts types bootstrapés.

L'échantillon bootstrapé est alors (x_i, y_i^*) sur lequel des régressions quantiles sont mises en œuvre pour obtenir la valeur bootstrap de β . Elle est solution de :

$$\hat{\beta}^*(\theta) = \underset{\beta(\theta)}{\text{Min}} \sum_i \rho_\theta(y_i^* - x_i \beta(\theta)).$$

B répétitions de cette procédure permettent d'obtenir B valeurs $\hat{\beta}(\theta_1), \dots, \hat{\beta}(\theta_B)$.

On a alors :

$\sqrt{n}(\hat{\beta}(\theta) - \beta(\theta)) \rightarrow N(0, \hat{\Lambda}(\theta))$, où $\hat{\Lambda}(\theta)$ est la matrice de variance-covariance asymptotique. Elle est égale à :

$$\hat{\Lambda}(\theta) = \frac{n}{B} \sum_{b=1}^B (\hat{\beta}(\theta_b) - \hat{\beta}(\theta)) (\hat{\beta}(\theta_b) - \hat{\beta}(\theta))' \quad (23)$$

Le Bootstrap par paires

Ce deuxième algorithme de bootstrap est utilisé lorsque les résidus sont indépendants, mais pas nécessairement identiquement distribués. Avec cette méthode, le rééchantillonnage ne porte plus sur les résidus, mais sur les paires $(x_i, y_i), i = 1, \dots, n$.

La méthode consiste à générer des couples de variables (x_i^*, y_i^*) tirés au hasard et avec remise dans l'échantillon initial pour constituer l'échantillon bootstrap.

La valeur bootstrap de β est obtenue en mettant en œuvre les régressions quantiles sur cet échantillon. $\hat{\beta}^*(\theta) = \underset{\beta(\theta)}{\text{Min}} \sum_i \rho_\theta(y_i^* - x_i^* \beta(\theta))$.

Cette procédure est répétée B fois pour obtenir ici encore B estimations de $\hat{\beta}^*(\theta)$.

La matrice de variance-covariance asymptotique obtenue avec cette deuxième méthode est, bien entendu, la même qu'en (23).

Remarque

En pratique, il est possible de choisir l'une ou l'autre méthode, qui sont équivalentes ; tout dépend de la « confiance » qu'on a dans le modèle initial. Le bootstrap par paires ne nécessite pas d'hypothèses sur le modèle et suppose uniquement que les paires sont des réalisations de la population. Étant moins sensible au modèle que le bootstrap des résidus, il sera préférable de l'utiliser en cas de doute sur la spécification du modèle.

L'estimation de cette matrice de variance-covariance permet donc de construire des intervalles de confiance sur les coefficients estimés par les régressions quantiles et de mettre en œuvre différents tests statistiques.

C. Les tests statistiques :

Plusieurs tests sont disponibles, et notamment le test de Wald de significativité des coefficients.

Le test de Wald (Koenker et Bassett, 1982-b)

Le test de Wald est un test de contraintes linéaires. Il prend la forme :

$$\begin{cases} H0 : R\beta(\theta) = r \\ HA : R\beta(\theta) \neq r \end{cases} \quad (24)$$

Dans le modèle $y_i = x_i' \beta(\theta) + u_i$, ce test consiste à se demander si, pour une variable explicative x_k , les coefficients estimés des différentes régressions quantiles sont significativement différents les uns des autres. Si c'est le cas, il permet de justifier l'utilisation des régressions quantiles plutôt que toute autre méthode d'estimation (les moindres carrés ordinaires par exemple)⁴⁸. C'est également un moyen de tester l'hétérogénéité des données.

La statistique de Wald est :

$$W = n(R\hat{\beta}(\theta) - r)'(R\hat{\Lambda}(\theta)R')^{-1}(R\hat{\beta}(\theta) - r) \quad (25)$$

où $\hat{\Lambda}(\theta)$ est la matrice de variance-covariance asymptotique du coefficient estimé par la régression quantile.

Si $p = \text{rg}(R)$, alors, sous $H0$, $W \rightarrow \chi_{(p)}$.

Le test du rapport de vraisemblance

La distribution de Laplace asymétrique (ou densité double exponentielle) est définie par :

$$f(u) = \theta(1 - \theta) \exp(-\rho_\theta(u)).$$

Koenker et Machado (1999) ont montré que l'estimateur de la régression quantile, $\hat{\beta}(\theta)$, peut aussi être considéré comme l'estimateur du quasi-maximum de vraisemblance, basé sur la distribution de Laplace asymétrique.

Dès lors, il est possible de mettre en œuvre le test du rapport de vraisemblance qui, comme le test de Wald, consiste à tester l'hypothèse nulle précédente contre l'hypothèse alternative. Si

⁴⁸ En effet, cela signifie que les pentes des différentes droites de régression estimées sont significativement différentes les unes des autres et donc que l'influence d'une variable x_k n'est pas identique en tous points de la distribution de Y . Dans le cas contraire, les régressions quantiles n'apportent aucune information supplémentaire par rapport à une régression par moindres carrés ordinaires car un seul coefficient est suffisant pour résumer l'influence de cette variable explicative sur la variable expliquée.

H_0 est vraie, alors imposer cette contrainte ne doit pas conduire à une importante diminution de la log-vraisemblance.

Le test est donc basé sur la différence de la log-vraisemblance entre le modèle non contraint et le modèle contraint.

Soit $V_{nc}(\theta) = \underset{\beta(\theta)}{\text{Min}} \sum_i \rho_\theta(y_i - x_i \beta(\theta))$ la fonction objectif évaluée dans le cas non contraint

(on note $\hat{\beta}(\theta)$ l'estimateur obtenu).

Et $V_c(\theta) = \underset{\beta(\theta) \text{ R}\beta(\theta)=r}{\text{Min}} \sum_i \rho_\theta(y_i - x_i \beta(\theta))$ la fonction objectif évaluée dans le cas contraint (on

note $\tilde{\beta}(\theta)$ l'estimateur obtenu).

Sous l'hypothèse de perturbations *iid* et en notant $\hat{s}_y(y)$ la fonction de rareté estimée⁴⁹, alors Koenker et Machado (1999) montrent que :

$$T_{LR}(\theta) = \frac{2(V_c(\beta) - V_{nc}(\theta))}{\theta(1-\theta)s_y(\theta)} \rightarrow \chi_q^2, \text{ où } q = \text{Rg}(R) \quad (26)$$

Ces deux tests (test de Wald et test du rapport de vraisemblance) sont asymptotiquement équivalents (Koenker et Machado, 1999).

7. Les propriétés des régressions quantiles

A. Équivariance

Les régressions quantiles ont des propriétés « d'équivariance » utiles : la reparamétrisation ou le changement d'échelle ne modifient pas les résultats des régressions quantiles.

Modifions les notations utilisées jusqu'ici et notons $\hat{\beta}(\theta, y, X)$ le coefficient estimé par la $\theta^{\text{ème}}$ régression quantile (pour tout $\theta \in [0,1]$) à partir de l'échantillon (y, X) . Koenker et Bassett (1978), théorème 3.2, montrent que les régressions quantiles possèdent de bonnes propriétés :

$$\begin{cases} \hat{\beta}(\theta, \lambda y, X) = \lambda \hat{\beta}(\theta, y, X), \lambda \in [0, +\infty] & (1) \\ \hat{\beta}(\theta, -\lambda y, X) = \lambda \hat{\beta}(1-\theta, y, X), \lambda \in [-\infty, 0] & (2) \\ \hat{\beta}(\theta, y + X\delta, X) = \hat{\beta}(\theta, y, X) + \delta, \delta \in \mathfrak{R}^K & (3) \\ \hat{\beta}(\theta, y, XA) = A^{-1} \hat{\beta}(\theta, y, X), A_{K \times K} \text{ singulière} & (4) \end{cases} \quad (27)$$

Illustrons l'ensemble de ces propriétés à l'aide d'exemples graphiques et numériques.

⁴⁹ Il s'agit ici de la fonction de rareté qui intervient dans la matrice de variance covariance et définie dans la section 6.1 (mais estimée par une méthode que nous n'avons pas présentée dans ce document).

Pour cela, nous simulons un échantillon de cinquante observations⁵⁰ dans lequel $x \rightarrow U(0;1)$, $\varepsilon \rightarrow N(0;1)$ et y est défini de telle sorte que $y = 3 + 2x + \varepsilon$.

Propriétés (1) et (2)

Elles impliquent que les coefficients des régressions quantiles sont invariants d'échelle.

Par la propriété (1), si y_i est multiplié par un facteur $\lambda > 0$ et devient $y_i^* = \lambda y_i$, alors la régression quantile de y_i^* sur x_i donne un nouvel estimateur qui change proportionnellement : ce facteur multiplicatif λ affecte à la fois la constante et la pente de l'équation à estimer.

Ainsi, $\hat{\alpha}(\theta)^* = \lambda \hat{\alpha}(\theta)$ et $\hat{\beta}(\theta)^* = \lambda \hat{\beta}(\theta)$, avec $\lambda > 0$.

En revanche, si $\lambda < 0$ (propriété (2)), alors $y_i^* = \lambda y_i$ change de signe donc les erreurs positives de la fonction objectif deviennent négatives et inversement.

Le nouvel estimateur basé sur y_i^* est alors $\hat{\alpha}(1-\theta)^* = \lambda \hat{\alpha}(\theta)$ pour la constante et $\hat{\beta}(1-\theta)^* = \lambda \hat{\beta}(\theta)$ pour la pente, avec $\lambda < 0$.

Exemples

Tableau 2 : Résultats des estimations : illustration des propriétés (1) et (2)

Régression	1	2	3
	RQ (0,25) sur y (Situation Initiale)	RQ (0,25) sur y^* (Propriété (1), avec $\lambda = 2$)	RQ (0,75) sur y^* (Propriété (2), avec $\lambda = -2$)
constante (écart-type)	2,403 (0,439)	4,806 (0,879)	- 4,806 (0,879)
pente (écart-type)	2,172 (0,822)	4,343 (1,644)	- 4,343 (1,644)
R ²	0,198	0,198	0,198

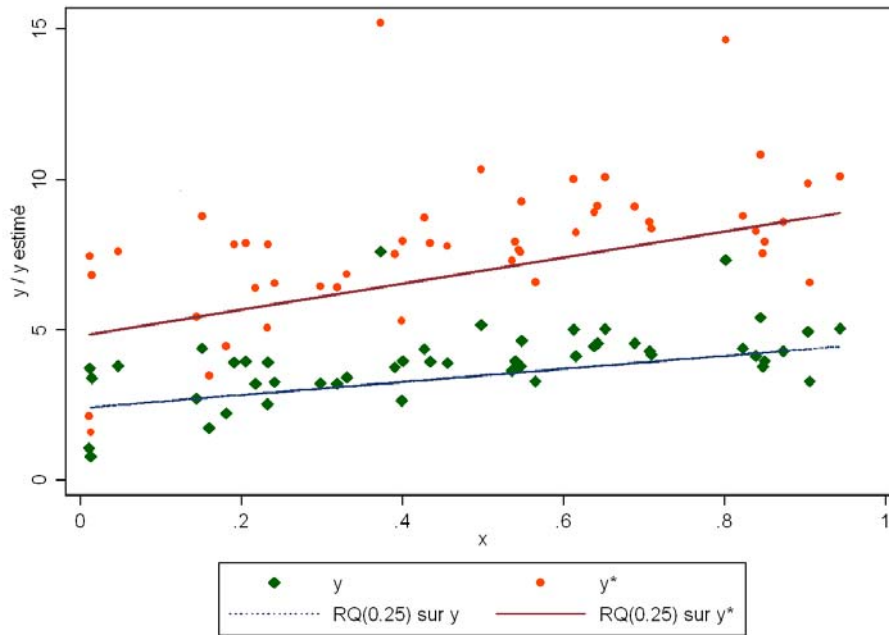
Les coefficients et les écarts-types obtenus dans la régression 2 sont 2 fois ceux de la régression 1.

Par ailleurs, les coefficients et les écarts-types obtenus dans la régression 3 sont (-2) fois ceux de la régression 1.

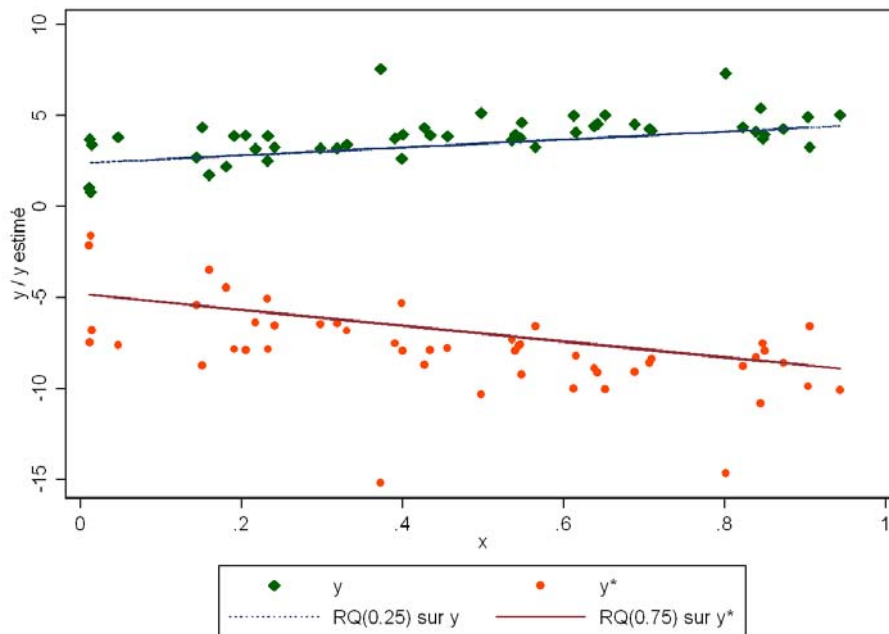
Graphiquement, il est possible d'observer les changements d'échelle :

⁵⁰ Prendre un échantillon plus grand ne modifie pas ces propriétés. Le travail sur un échantillon de petite taille permet ici une meilleure visualisation des transformations des données et n'a qu'un objectif pédagogique.

Graphique 6 : Illustration de la propriété (1)



Graphique 7 : Illustration de la propriété (2)



Propriété (3)

Une nouvelle équation est considérée : $y_i^* = y_i + X\gamma$. Dans ce cas, l'estimation de la pente est modifiée, mais pas celle de la constante.

Alors $\hat{\alpha}(\theta)^* = \hat{\alpha}(\theta)$ mais $\hat{\beta}(\theta)^* = \hat{\beta}(\theta) + \gamma$.

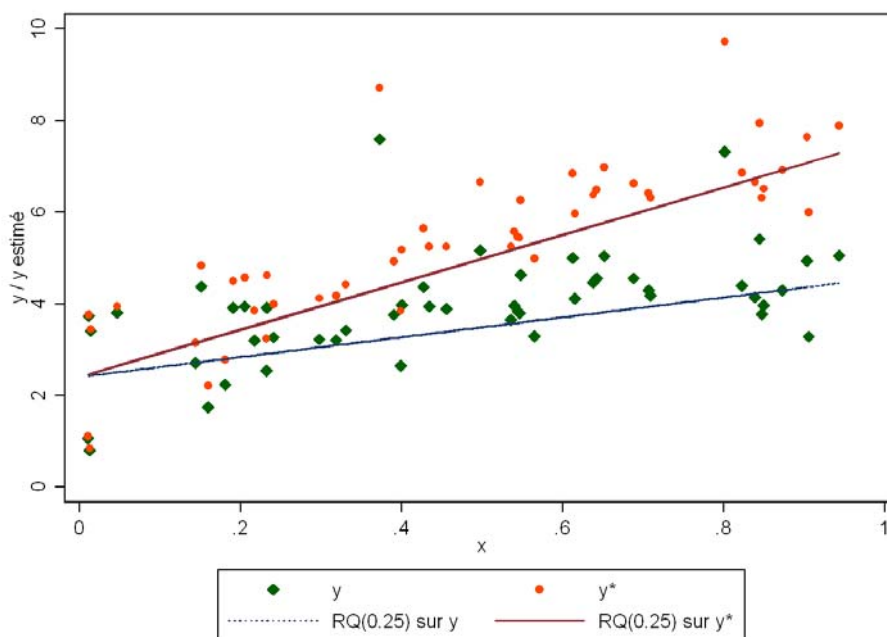
Exemple

Tableau 3 : Résultats des estimations : illustration de la propriété (3)

Régression	1	2
	RQ (0,25) sur y (Situation Initiale)	RQ (0,25) sur y^* (Propriété (3), avec $y_i^* = y_i + 3x_i$)
constante (écart-type)	2,403 (0,439)	2,403 (0,439)
pente (écart-type)	2,172 (0,822)	5,172 (0,439)
R^2	0,198	0,51

La constante reste identique, mais la pente est modifiée (égale à la pente de la régression (1) +3). Les écarts types restent toutefois identiques (la dispersion globale des observations n'a en effet pas varié par rapport à la situation initiale, puisque le changement d'échelle touche toutes les observations identiquement).

Graphique 8 : Illustration de la propriété (3)



Propriété (4)

Si on « reparamétrise » X , $\hat{\beta}(\theta)$ est lui aussi reparamétrisé par l'inverse de la transformation apportée à X .

Remarque

Les moindres carrés ordinaires vérifient également l'ensemble les propriétés (1), (2) et (4).

B. Invariance aux transformations monotones :

Les régressions quantiles, parce qu'elles ont comme origine l'ordonnement des observations, sont invariantes à une transformation (linéaire ou non) monotone de y .

Ainsi, pour toute fonction $h(\cdot)$ monotone, non décroissante sur \mathfrak{R} on a, (Powell, 1989) :

$$\boxed{Q_{h(Y)}(\theta|x) = h(Q_Y|x)} \quad (28)$$

En effet, $\Pr(Y < y|x) = \Pr(h(Y) < h(y)|x)$: les quantiles de la variable transformée $h(Y)$ sont identiques aux quantiles de la variable initiale Y .

Pour estimer Y en utilisant les régressions quantiles, il est donc tout à fait possible d'estimer $\log(Y)$, puis de transformer de nouveau les données (en passant à l'exponentielle) pour retourner à l'échelle initiale, sans perte d'information et sans modifier l'estimation des quantiles de Y conditionnels à x ⁵¹.

Si $h(Y) = \log(Y)$, alors :

$$Q_{\log(Y)}(\theta|x) = \log(Q_Y(\theta|x)) = x_i' \beta(\theta)$$

$$\text{Et donc } Q_Y(\theta|x) = \exp(x_i' \beta(\theta))$$

Cette propriété ne se vérifie pas dans le cadre des moindres carrés ordinaires. En effet, en général, $E(h(Y)) \neq h(E(Y))$, sauf si $h(\cdot)$ est linéaire.

8. Les avantages des régressions quantiles par rapport aux régressions par moindres carrés ordinaires

A. Une méthode semi-paramétrique

Les régressions quantiles constituent un intermédiaire entre les méthodes paramétriques et non paramétriques.

En effet, aucune forme paramétrique (loi de poisson, loi normale, loi binomiale,...) n'est imposée à la loi des perturbations. En revanche, la forme fonctionnelle reste paramétrique (on définit par exemple une relation linéaire entre X et Y de type $Y = a + bX$).

⁵¹ Évidemment, les régressions quantiles ne dispensent pas de s'interroger sur la forme de la relation entre la variable expliquée et les variables explicatives, puisque si $Q_{\log(Y)}(\theta|x) = \log(Q_Y(\theta|x))$, en revanche $Q_{\log(Y)}(\theta|x) \neq Q_{\log(Y)}(\theta|\log(x))$.

B. Une estimation de la distribution globale de Y :

Les moindres carrés ordinaires, tout comme la régression médiane, n'estiment que la tendance centrale d'une distribution (la moyenne ou la médiane conditionnelle). En revanche, les régressions quantiles permettent d'estimer l'ensemble de la distribution conditionnelle.

Si on considère que $E(y|x) = \int Q_Y(\theta|x)d\theta$, alors le modèle linéaire traditionnel estimé par les moindres carrés ordinaires peut être considéré comme une « somme » des effets aux différents quantiles de la distribution.

Mais cette agrégation par la moyenne peut masquer certains effets.

L'obtention de $\hat{\beta}_k = 0$ par l'estimation par moindres carrés ordinaires (effet non significatif du coefficient associé à la variable explicative x_k), peut conduire à deux conclusions distinctes :

- D'une part, en moyenne, la variable x_k n'a pas d'influence sur la variable expliquée et n'affecte donc pas la distribution de y .
- D'autre part, les effets à différents points de la distribution conditionnelle de Y se compensent (un effet positif pour le haut de la distribution et un effet négatif pour le bas par exemple). En moyenne, l'effet mesuré est donc nul, bien qu'il existe une relation entre x et y .

Les régressions quantiles permettent de mesurer l'influence de la variable explicative à chaque quantile de la distribution conditionnelle de y et d'indiquer comment les variables explicatives affectent différentes parties de la distribution conditionnelle de Y .

C. La robustesse aux « outliers » (valeurs atypiques) :

Les régressions quantiles ont en premier lieu été développées afin de constituer une alternative aux moindres carrés ordinaires plus robuste aux outliers.

Par la forme même de la fonction objectif à minimiser pour obtenir les quantiles de la régression (c'est une somme pondérée d'écarts absolus), les régressions quantiles sont peu sensibles aux outliers, ces derniers recevant moins de poids que dans le cadre de l'estimation par moindres carrés ordinaires (où ils sont élevés au carré).

Par ailleurs, considérons une solution $\hat{\beta}(\theta)$ obtenue à partir de l'échantillon (y, X) .

Si certaines observations sont modifiées, de telle sorte que $y_i^* = y_i + \Delta y_i$, alors la condition d'optimalité obtenue reste identique si le passage de y_i à y_i^* ne modifie pas le signe des résidus.

En effet, dans la détermination de la solution $\hat{\beta}(\theta)$, seul compte le signe des résidus (et notamment le nombre de résidus positifs et négatifs), mais leur valeur importe peu dans l'estimation des quantiles conditionnels.

La nouvelle valeur de y_i , notée y_i^* n'influence donc l'estimation que dans la mesure où elle est au-dessus ou au-dessous de l'hyperplan estimé, mais la distance à cet hyperplan estimé importe peu.

En outre, si toutes les observations de l'échantillon sont utilisées pour déterminer lesquelles correspondent aux différents quantiles de l'échantillon, seuls quelques points sont ensuite utilisés pour estimer le $\theta^{\text{ème}}$ quantile conditionnel. Tant qu'on ne modifie pas la valeur de ces observations et que le nombre de résidus positifs et négatifs reste identique, les estimations sont inchangées.

Notons toutefois que si les régressions quantiles sont robustes aux outliers de la variable expliquée y , elles le sont beaucoup moins aux contaminations de la variable explicative x .

Quelques éléments d'interprétation

Reprenons le cas initial de l'estimation de la médiane non conditionnelle d'un échantillon, qui pourra ensuite être généralisé, afin de comprendre dans quelle mesure les régressions quantiles sont insensibles aux valeurs atypiques d'un échantillon (Y_1, \dots, Y_n) .

Soit q_m , la médiane de l'échantillon obtenue dans la situation initiale, c'est-à-dire à partir de l'échantillon $(Y_1, \dots, Y_i, \dots, Y_n)$. Elle est telle que $N_1 = N_2$: il y a autant d'observations au-dessus qu'au-dessous de q_m .

Supposons qu'une des valeurs de l'échantillon, par exemple Y_i , soit modifiée. L'échantillon devient alors $(Y_1, \dots, Y_i', \dots, Y_n)$. 2 cas sont possibles :

- soit Y_i' reste inférieure (resp. supérieure) à la médiane, comme Y_i
- soit Y_i' devient supérieure (resp. inférieure) à Y_i alors qu'elle lui était inférieure (supérieure)

- Cas 1 :

Supposons que Y_i soit modifiée de telle sorte que Y_i' reste inférieure à la médiane obtenue dans la situation initiale. Même si la valeur de Y_i' est beaucoup plus faible que celle de Y_i , il reste toujours N_1 observations supérieures à la médiane et N_2 observations inférieures à la médiane. La médiane q_m qui est la valeur de l'échantillon telle que $N_1 = N_2$ n'est donc pas modifiée.

- Cas 2 :

Supposons que Y_i soit modifiée de telle sorte que Y_i' devienne supérieure à la médiane, alors qu'elle lui était inférieure dans la situation initiale.

Le nombre d'observations inférieures et supérieures à la médiane initiale est donc modifié : il y a maintenant $N_1' = N_1 + 1$ observations supérieures à la médiane initiale et $N_2' = N_2 - 1$ observations qui lui sont inférieures.

La nouvelle médiane de l'échantillon est maintenant telle que $N_1' = N_2'$, soit :

$$N_1' = N_2'$$

$$\Leftrightarrow N_1 + 1 = N_2 - 1$$

$$\Leftrightarrow N_1 + 2 = N_2$$

q_m n'est donc plus solution de ce nouvel échantillon : la modification de l'ordre des observations a transformé la solution du programme d'optimisation.

Dans le cadre des régressions quantiles, il est possible de raisonner de manière semblable. Seul le *signe* des résidus compte pour la détermination de la solution. Tant que le nombre de résidus positifs et négatifs n'est pas modifié, la solution reste identique.

Exemples

Soit un échantillon de cinquante observations dans lequel $x \rightarrow U(0;1)$, $\varepsilon \rightarrow N(0;1)$ et $y = 3 + 2x + \varepsilon$ ⁵².

Il s'agit ici d'un modèle linéaire simple dans lequel les hypothèses des moindres carrés ordinaires sont vérifiées : les perturbations sont d'espérance nulle et homoscédastiques, la variable explicative est certaine et de variance non nulle. De plus, y est défini comme une fonction linéaire croissante de x .

Les moindres carrés ordinaires semblent donc être la méthode d'estimation conseillée pour ce modèle.

Toutefois, l'exemple suivant montre que ces derniers sont beaucoup plus sensibles aux valeurs extrêmes de la variable y (et dans une moindre mesure de la variable x) que la régression médiane⁵³.

➤ *Les Outliers de la variable dépendante*

Les résultats des estimations de trois situations sont comparés :

- La situation initiale
- Le Cas n° 1 : la variable y est modifiée mais y_i' reste du même côté de la droite de régression estimée initiale.
- Le Cas n° 2 : la variable y est modifiée mais y_i' passe au-dessus de droite de régression estimée initiale.

⁵² Comme précédemment, prendre un échantillon de taille plus importante ne modifie pas la propriété.

⁵³ Seules les régressions médianes sont ici présentées, afin de comparer plus facilement les propriétés des régressions quantiles avec celles des moindres carrés ordinaires. Toutefois, ces propriétés sont généralisables à l'ensemble des régressions quantiles.

Tableau 4 : Résultats des estimations : régression par moindres carrés ordinaires et régression médiane dans les trois situations

	Moindres carrés ordinaires			Régression médiane		
	MCO (situation initiale)	MCO (outlier sur y, cas n° 1)	MCO (outlier sur y, cas n° 2)	MEA (situation initiale)	MEA (outlier sur y, cas n° 1)	MEA (outlier sur y, cas n° 2)
Constante (écart type)	2,816 (0,285)	3,134 (0,533)	2,804 (0,330)	2,926 (0,259)	2,926 (0,259)	2,887 (0,271)
Pente (écart type)	2,337 (0,523)	2,13 (0,978)	2,150 (0,606)	1,934 (0,470)	1,934 (0,470)	1,99 (0,493)
R ²	0,293	0,09	0,210	0,137	0,107	0,119

L'estimation par moindres carrés ordinaires est très sensible à cette valeur atypique, que ce soit dans le cas n° 1 ou le cas n° 2. Les coefficients estimés varient fortement, les écarts-types augmentent et le R² diminue. La nouvelle droite de régression estimée est très éloignée de la droite de régression initiale (graphique 9 pour le cas n° 1).

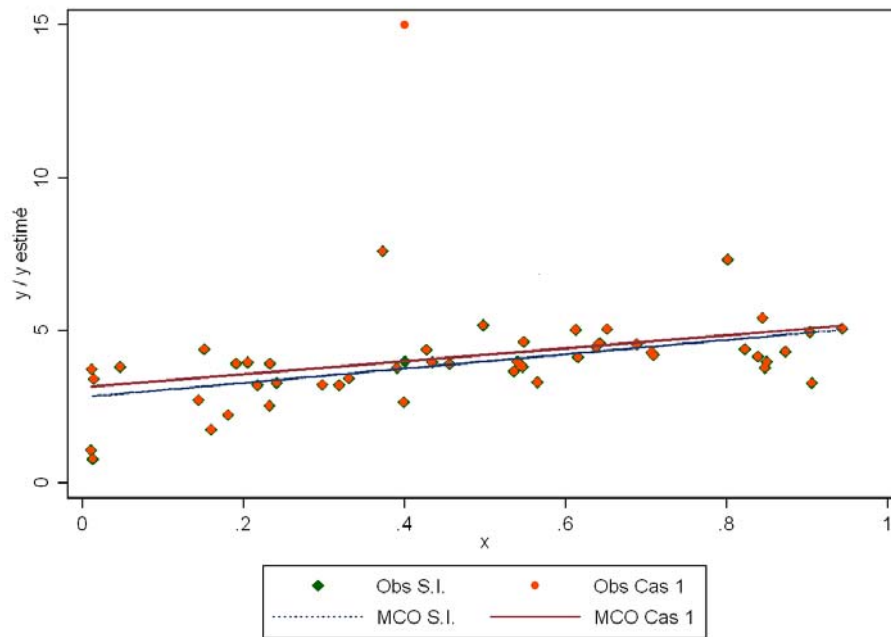
La régression médiane est insensible à l'outlier sur y dans le cas n° 1 : les coefficients estimés sont identiques et les droites de régression estimées sont donc confondues (graphique 10). En effet, l'observation modifiée reste du même côté de la droite de régression initiale ; le nombre de résidus positifs et négatifs étant inchangé, l'estimation l'est aussi.

Toutefois, les écarts types augmentent car la valeur atypique introduite dans les données diminue la concentration initiale des données. La valeur atypique n'influence donc pas les résultats des estimations au premier ordre mais a une influence au second ordre.

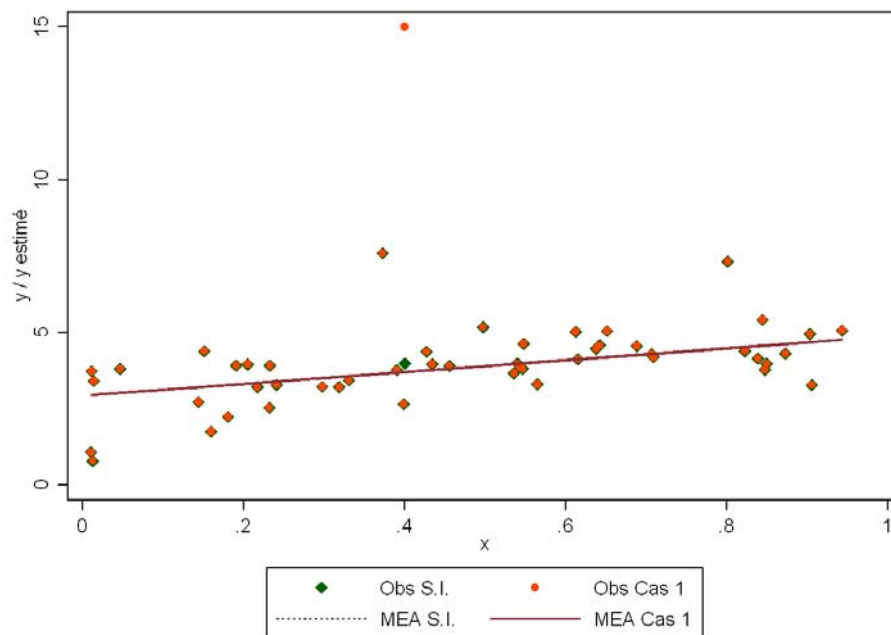
La régression médiane est en revanche sensible à l'outlier sur y dans le cas n° 2 : les coefficients estimés varient, les droites de régression estimées sont différentes (graphique 11). L'observation modifiée passe en effet de l'autre côté de la droite de régression initiale. Le nombre de résidus positifs diminue et le nombre de résidus négatifs augmente : cela modifie la position de la nouvelle droite de régression.

Par ailleurs, les écarts types sont également modifiés. L'outlier, a une influence sur les résultats des estimations au premier et au second ordre.

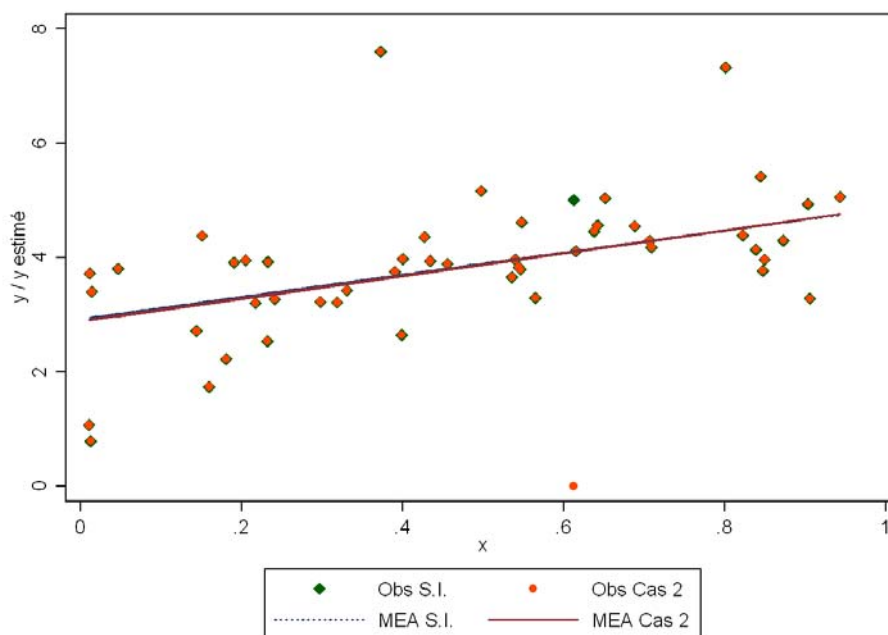
Graphique 9 : Influence d'une valeur atypique de y sur les estimations par moindres carrés ordinaires



Graphique 10 : Influence d'une valeur atypique de y (Cas n° 1) sur les estimations par la régression médiane



Graphique 11 : Influence d'une valeur atypique de y (Cas n° 2) sur les estimations par la régression médiane



➤ *Les Outliers de la variable expliquée*

Si les régressions quantiles sont robustes aux valeurs atypiques sur la variable dépendante, elles sont en revanche, comme les moindres carrés ordinaires, beaucoup plus sensibles aux outliers sur la ou les variables explicatives.

D. Les régressions quantiles plutôt que les moindres carrés ordinaires ou les régressions robustes

En présence de valeurs atypiques, il est possible de corriger les moindres carrés ordinaires en utilisant des moindres carrés pondérés (MCP). Comparons les estimations obtenues en présence de données atypiques avec les MCP et les régressions quantiles.

Remarque : l'estimation par moindres carrés pondérés

Soit le modèle $y_i = x_i' \beta + u_i$.

La méthode des moindres carrés pondérés consiste à estimer ce modèle par moindres carrés ordinaires, pour en déduire la valeur des résidus estimés pour chaque observation : $\hat{u}_i = y_i - \hat{y}_i$.

Un poids w_i est ensuite assigné à chaque individu, ces poids devant respecter certaines conditions :

- $\sum_{i=1}^n w_i = 1$
- le poids w_i assigné à chaque individu doit être d'autant plus faible que le résidu est grand.
- Les observations les plus atypiques (pour lesquelles la distance de Cook est supérieure à 1) ne sont pas incluses dans l'estimation⁵⁴.

Enfin, le modèle est de nouveau estimé, en incluant les poids pour chaque observation.

Pour estimer un modèle du type $y = a + bx_1 + cx_2 + u$, la procédure sous STATA est la suivante⁵⁵ :

rreg y x1 x2, genwt(w)

⇒ l'option **genwt(w)** permet de conserver les poids qui sont générés par la procédure.

Comparaison des différentes estimations

Reprenons l'échantillon de données précédent et incluons quatre valeurs atypiques sur la variable à expliquer.

L'objectif est de comparer l'estimation de la tendance centrale de l'échantillon en utilisant trois méthodes distinctes :

- les moindres carrés ordinaires (MCO) estiment la moyenne conditionnelle de y
- La régression robuste (RR) estime aussi la moyenne conditionnelle de y , mais en corrigeant l'influence des outliers
- la régression médiane (MEA) estime la médiane conditionnelle de y

Tableau 5 : Résultats des estimations par les trois méthodes

	MCO	RR	MEA
Constante (écart type)	3,319 (0,505)	2,766 (0,234)	2,887 (0,271)
Pente (écart type)	1,267 (0,926)	2,155 (0,430)	1,990 (0,493)

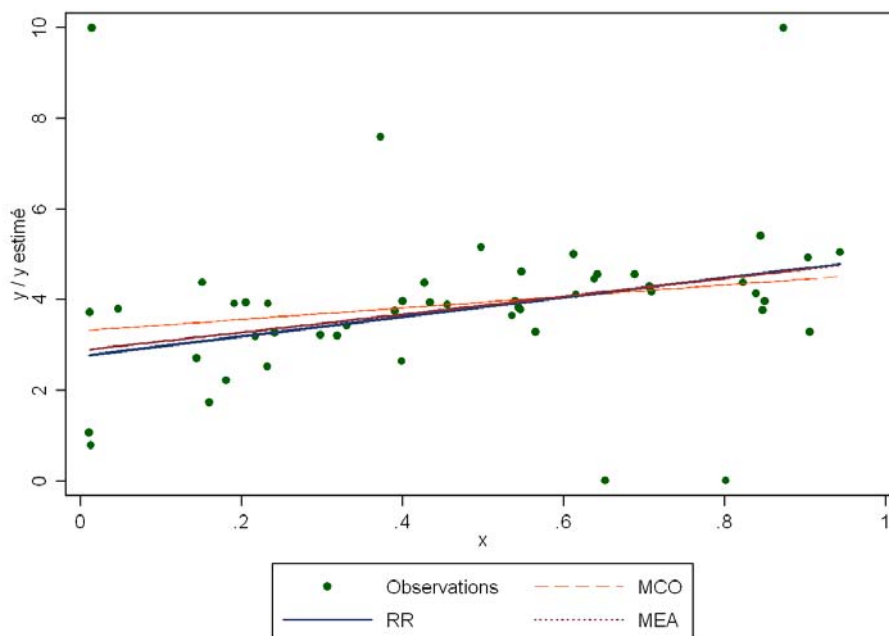
⁵⁴ La « distance de Cook » mesure l'influence d'une observation sur l'ensemble des estimations. Elle consiste à comparer un paramètre estimé sans la $i^{\text{ème}}$ observation avec le même paramètre estimé à partir de toutes les observations. Elle est définie

ainsi : $D_i = \frac{\sum_{j=1}^n (\hat{y}_{(i)j} - \hat{y}_j)^2}{2\hat{\sigma}^2}$, $i = 1 \dots n$. Les valeurs atypiques sont les observations pour lesquelles $D_i > 1$. De tels

points peuvent exercer un « effet de levier » et fausser la régression, il est préférable de ne pas les inclure dans la régression.

⁵⁵ La procédure d'estimation des régressions quantiles sous STATA n'est pas présentée ici car elle est décrite de façon détaillée dans la section 11 de ce document.

Graphique 12 : Comparaison de la robustesse des moindres carrés ordinaires, de la régression robuste et de la régression médiane, en présence de deux observations atypiques



Comme les régressions quantiles, les régressions robustes ne sont pas sensibles aux outliers, alors que les moindres carrés ordinaires le sont fortement.

Les régressions robustes et médiane donnent des résultats assez proches. Toutefois, les écarts types sont plus faibles avec la régression robuste. En effet, à la différence de la régression médiane qui prend en compte les outliers dans l'estimation, la régression robuste ignore simplement les valeurs les plus extrêmes (en leur accordant un poids nul), ce qui réduit la dispersion des données.

Les régressions robustes ne constituent qu'une correction par rapport à l'estimation par moindres carrés ordinaires, et même si ces régressions sont robustes aux outliers, elles ne permettent que d'estimer la tendance centrale d'une distribution (la moyenne) et pas ses différents quantiles. L'intérêt de l'utilisation des régressions quantiles reste donc toujours présent.

E. Les régressions quantiles permettent de détecter l'hétéroscédasticité des données

Les MCO supposent que la distribution conditionnelle des honoraires est homogène : la relation entre les variables X et Y est identique, quelle que soit la place à laquelle on se situe dans la distribution conditionnelle de Y (un seul nombre synthétise toute cette information).

Les régressions quantiles ne font aucune hypothèse sur l'homogénéité de la distribution conditionnelle et acceptent même l'hypothèse d'hétéroscédasticité. En outre, il est possible d'utiliser les régressions quantiles pour montrer qu'une distribution conditionnelle n'est pas homogène.

En effet, si la distribution conditionnelle est homogène, les coefficients des variables explicatives estimés aux différents quantiles de la distribution de Y sont identiques, signifiant une influence de X sur Y identique, à tous les quantiles.

Un exemple permet d'illustrer cette propriété. Deux échantillons de trois cent observations sont simulés :

- Le premier échantillon (noté 1) est homoscédastique : $X \rightarrow U(0;1)$, $\varepsilon \rightarrow N(0;1)$ et y_1 est tel que $y_1 = 3 + 2x + \varepsilon$.
- Le deuxième échantillon (noté 2) est hétéroscédastique : $X \rightarrow U(0;1)$, $\nu \rightarrow N(0;9x^2)$ et y_2 est tel que $y_2 = 3 + 2x + \nu$.

Pour les deux échantillons, cinq régressions quantiles sont estimées, associées à $\theta = \{0.1, 0.25, 0.5, 0.75, 0.9\}$

➤ *Échantillon 1*

Tableau 6 : Résultats des estimations dans le cas homoscédastique

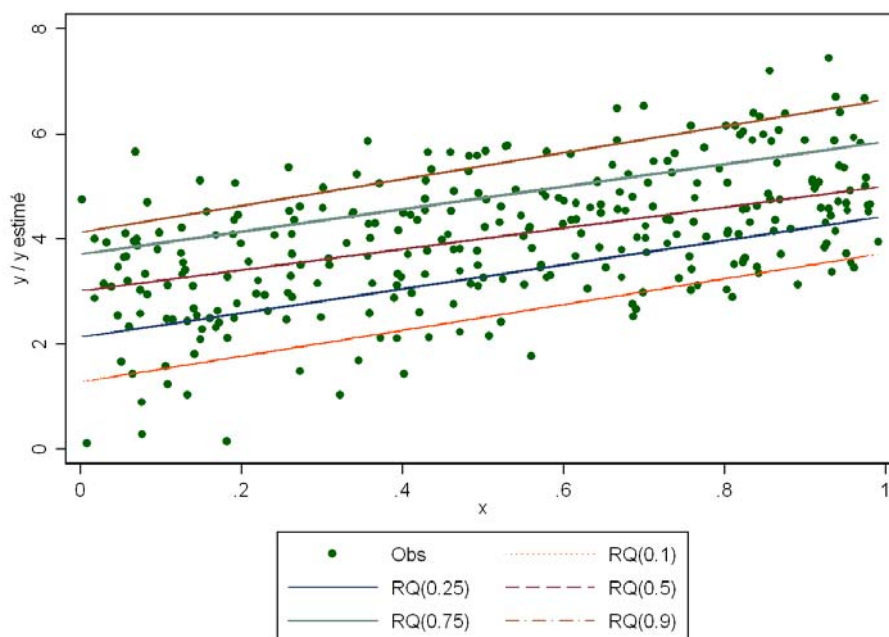
	RQ(0,1)	RQ(0,25)	RQ(0,5)	RQ(0,75)	RQ(0,9)
constante (écart type)	1,273 (0,226)	2,129 (0,155)	3,011 (0,156)	3,714 (0,177)	4,132 (0,188)
pente (écart type)	2,457 (0,376)	2,307 (0,261)	1,997 (0,263)	2,142 (0,299)	2,524 (0,318)
R ²	0,175	0,145	0,134	0,129	0,161

Les pentes ne sont pas significativement différentes entre les quantiles⁵⁶, seules les constantes diffèrent, et ceci, mécaniquement, du fait de la variation des données de l'échantillon.

Le graphique des valeurs prédites de Y en fonction de la variable X est constitué de droites parallèles.

⁵⁶ C'est la mise en œuvre du test de Wald, dont la procédure est détaillée dans la section 7, qui permet d'aboutir à cette conclusion.

Graphique 19 : Comparaison de cinq régressions quantiles dans le cas homoscédastique :



➤ *Échantillon 2*

Tableau 7 : Résultats des estimations dans le cas hétéroscédastique

	RQ(0,1)	RQ(0,25)	RQ(0,5)	RQ(0,75)	RQ(0,9)
constante (écart type)	2,926 (0,241)	2,947 (0,172)	3,00 (0,141)	2,957 (0,217)	2,910 (0,150)
pente (écart type)	-1,937 (0,401)	0,082 (0,289)	2,026 (0,237)	4,484 (0,214)	6,694 (0,253)
R ²	0,053	0,002	0,084	0,205	0,341

Dans ce deuxième échantillon, les pentes sont significativement différentes entre les quantiles, les queues de distribution et la tendance centrale de l'échantillon varient différemment avec X . L'influence de cette variable est croissante : elle est négative au premier décile de la distribution de Y , puis croît régulièrement avec les quantiles de Y .

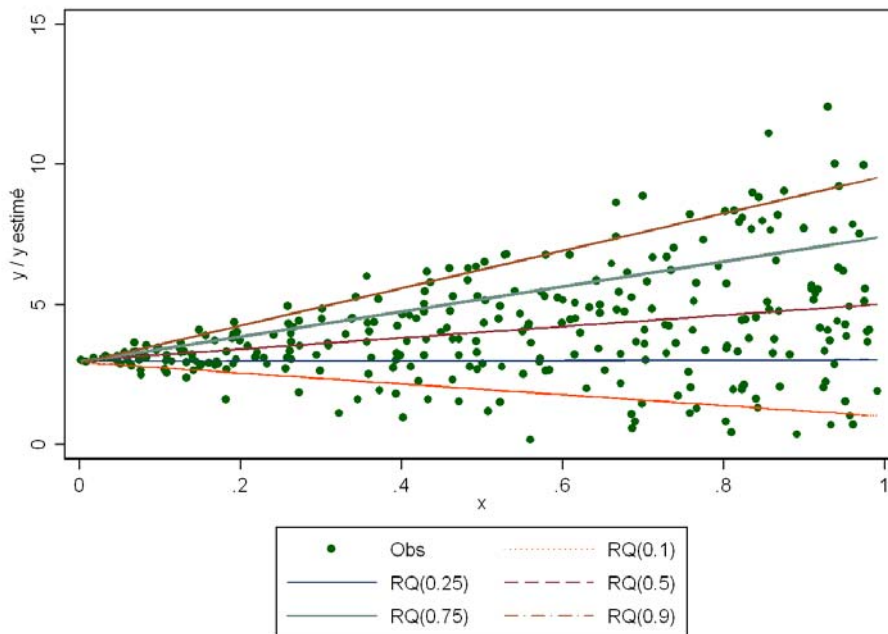
En revanche, les constantes, qui donnent la valeur du quantile de Y pour une valeur de X nulle, ne sont pas significativement différentes entre les quantiles. Ce n'est généralement pas le cas car cela sous-entendrait que les quantiles sont identiques. Ce phénomène est ici lié à la constitution de l'échantillon, puisque nous avons voulu mettre en avant un cas extrême d'hétéroscédasticité.

Signe de cette hétéroscédasticité des données, les différentes droites estimées ne sont plus parallèles les unes aux autres.

Remarque

Avec les régressions quantiles, l'hétéroscédasticité des données ne se voit pas dans les variances, comme c'est le cas avec l'estimation par moindres carrés ordinaires, mais sur les coefficients.

Graphique 14 : Comparaison de cinq régressions quantiles dans le cas hétéroscédastique



9. Mise en œuvre des régressions quantiles grâce au logiciel STATA

Plusieurs logiciels permettent de mettre en œuvre les régressions quantiles (R et STATA notamment). La version 9 de SAS a également intégré une version expérimentale des régressions quantiles.

L'exemple suivant permet de comprendre la mise en œuvre des régressions quantiles et leur interprétation grâce au logiciel STATA.

Il s'appuie sur les données du panel d'omnipraticiens libéraux relatives à l'année 2000, utilisé dans la première partie.

- La variable expliquée est le logarithme des honoraires.
- Les variables explicatives sont :
 - Le sexe de l'omnipraticien
 - Son ancienneté et ancienneté au carré
 - La densité d'omnipraticiens libéraux dans son département d'exercice

A. Les commandes utilisées pour la régression médiane

- La commande permettant de mettre en œuvre la régression médiane est la commande **qreg**.

```
• qreg lhono sexe anc anc2 dens
```

- Il est possible d'effectuer cette régression médiane en estimant des écarts types bootstrapés. La commande est alors **bsqreg**.

```
• bsqreg lhono sexe anc anc2 dens, reps(1000)
```

Remarque

reps(1000) indique le nombre d'itérations bootstrap à effectuer. Par défaut, STATA en réalise 20, mais ce nombre d'itérations est très insuffisant pour que les écarts types obtenus soient robustes.

B. Les commandes utilisées pour l'ensemble des régressions quantiles

- La régression médiane étant un cas particulier des régressions quantiles, ces dernières peuvent être estimées en utilisant les mêmes commandes que la régression médiane : **qreg** (pour la régression sans itérations bootstrap) ou **bsqreg** (pour la régression avec écarts types bootstrapés).

Une option doit cependant être rajoutée pour spécifier le quantile à estimer (toutes les valeurs $\theta \in [0,1]$ ⁵⁷) : **quant()**.

- ```
• Estimation de la régression du 1er quartile
qreg lhono sexe anc anc2 dens, quant(0.25)
• Estimation de la régression du 9ème décile, avec écarts types bootstrapés
qreg lhono sexe anc anc2 dens, quant(0.9) reps(1000)
```

- La commande **sqreg** permet d'estimer simultanément différentes régressions quantiles (avec des écarts types bootstrapés ou non).

- ```
• Estimation de plusieurs régressions quantiles  
sqreg lhono sexe anc anc2 dens, quant(0.1 0.25 0.5 0.75 0.9)  
• Estimation de plusieurs régressions quantiles, avec écarts types bootstrapés  
sqreg lhono sexe anc anc2 dens, quant(0.1 0.25 0.5 0.75 0.9) reps(1000)
```

⁵⁷ En pratique, les RQ(0) et RQ(1) ne sont pas mises en œuvre.

Les commandes utilisées pour les régressions inter-quantiles

- La commande **iqreg** permet d'estimer des intervalles inter-quantiles (avec des écarts types bootstrapés ou non).

- Régression inter-quartile
iqreg lhono sexe anc anc2 dens,quant(0.25 0.75)
- Régression inter-décile, écarts types bootstrapés
iqreg lhono sexe anc anc2 dens,quant(0.1 0.9)

C. Interprétation des résultats obtenus sous STATA pour le cas particulier de la régression médiane

Les résultats obtenus pour la régression médiane sont les suivants :

```

bsqreg lhono sexe anc anc2 dens, reps(1000)
(fitting base model)
(bootstrapping)

Median regression, bootstrap(1000) SEs
Number of obs =      5350 (1)
Raw sum of deviations 2145.854 (about 11.55995) (2)
Min sum of deviations 1993.256 (3)
Pseudo R2 =      0.0711 (4)

-----+-----
lhono | Coef. Std. Err. t P>|t| [95 % Conf. Interval]
-----+-----
sexe | -.3575396 .0164607 -21.72 0.000  -.3898093  -.3252699
anc | .0108141 .0032168  3.36 0.001  .0045078  .0171204
anc2 | -.00026 .0000869 -2.99 0.003  -.0004304  -.0000897
dens | -.004125 .000425 -9.71 0.000  -.0049582  -.0032918
_cons | 12.32738 .0555916 221.75 0.000  12.2184  12.43636
-----+-----

```

(1) donne le nombre d'observations utilisées pour l'estimation

(2) « **about** » donne la valeur du quantile non conditionnel, ici la médiane non conditionnelle : la médiane du logarithme des honoraires est environ 11,56.

« **raw sum of absolute deviations** » est la valeur de $\sum_i |y_i - 11,56| = 2145,8$

(3) donne la valeur de $\sum_i |y_i - \hat{Q}_Y(0,5)|$, c'est-à-dire la somme des écarts absolus des y_i aux $\hat{Q}_Y(0,5)$ estimés.

(4) C'est le pseudo R², avec $(4) = 1 - \frac{(3)}{(2)}$.

Le tableau donne ensuite la valeur de chaque coefficient estimé, le T de Student, l'écart type et l'intervalle de confiance associés.

Les résultats relatifs à l'estimation des autres régressions quantiles ne sont pas décrits ici, mais leur interprétation des résultats est identique.

VII. Bibliographie

1. Étude de la dispersion des honoraires des omnipraticiens :

- AUDRIC, S. 2006. « Analyse des carrières des médecins libéraux à partir de données de panels », DREES, *Document de Travail*, Série « Statistiques », n° 96, mai.
- BESSIERE, S. 2005. « La féminisation des professions de santé : données de cadrage », dans la *Revue Française des Affaires Sociales* (« Dynamiques professionnelles dans le champ de la santé »), n° 1, janvier-mars 2005, la Documentation Française.
- BEUDAERT, M. 1999. « Les honoraires des médecins généralistes entre 1985 et 1995 », DREES, *Études et Résultats* n° 15, avril.
- BILLAUT, A., 2006. « Les affectations en troisième cycle des études médicales en 2005 suite aux épreuves classantes nationales », DREES, *Études et Résultats* n°474, mars
- CNAMTS, 1984. « Le secteur libéral des professions de santé en 1983 », *Carnets statistiques* n° 12.
- CNAMTS, 1993. « Le secteur libéral des professions de santé en 1992 », *Carnets statistiques* n° 74.
- CNAMTS, 2002. « Le secteur libéral des professions de santé en 2000 », *Carnets statistiques* n° 108.
- CNAMTS, 2003. « Le secteur libéral des professions de santé en 2001 », *Carnets statistiques* n° 109.
- CNAMTS, 2003. « L'évolution des honoraires médicaux depuis 1980 », *Point de conjoncture* n° 16-17, septembre.
- DELATTRE, E. et DORMONT, B., 2000. « Induction de la demande de soins par les médecins libéraux français », *Économie et Prévision*, n° 142.
- DELATTRE, E. et DORMONT, B., 2005. « La régulation de la médecine ambulatoire en France : quel effet sur le comportement des médecins libéraux ? », DREES, *Solidarité et Santé* n° 1, janvier-mars.
- FIVAZ, C., GARIEPUY, S. et LE LAIDIER, S., 2000. « Disparités dans les évolutions des honoraires médicaux », CNAMTS, *Point Stat* n° 30, décembre
- FIVAZ, C. et LE LAIDIER, S., 2001. « Une semaine d'activité des généralistes libéraux », CNAMTS, *Point Stat* n° 33, juin.
- HARDY-DUBERNET, A-C. 2005. « Femmes en médecine : vers un nouveau partage des professions ? », dans la *Revue Française des Affaires Sociales* (« Dynamiques professionnelles dans le champ de la santé »), n° 1, janvier-mars 2005, la Documentation Française.
- HARDY-DUBERNET, A-C.(dir.), ARLIAUD, M., HORELLOU-LAFARGE, C., LE ROY, F et BLANC, M-A., (2001). « La réforme de l'internat de médecine de 1982 et ses effets sur les choix professionnels des médecins », *Rapport de recherche pour le programme INSERM/CNRS/MiRe98 : Processus de décision et changements des systèmes de santé*, novembre.
- LEGENDRE, N., 2005. « L'évolution sur dix ans des revenus libéraux des médecins, 1993-2003 », DREES, *Études et Résultats* n° 412, juillet
- LEGENDRE, N., 2006. « Les revenus libéraux des médecins en 2003 et 2004 », DREES, *Études et Résultats* n° 457, janvier.
- LUCAS-GABRIELLI, V. et SOURTY-LE GUELLEC, M.J., 2004. « Évolution de la carrière des médecins généralistes selon leur date d'installation (1979-2001) », *Questions d'économie de la santé* n° 81, CREDES, avril.

- NIEL, X. et PERRET, J-P., 2002. « Féminisation et Vieillesse des médecins au cours des années quatre-vingt-dix », *Données sociales 2002-2003*, pp. 385.
- SAMSON, A-L., 2006. « La dispersion des honoraires des omnipraticiens – analyse sur la période 1983-2003 », DREES, *Études et Résultats* n° 482, avril.

2. La méthode des régressions quantiles :

- BASSETT, G. W. et KOENKER, R. 1983. « An Empirical Quantile Function For Linear Models with iid Errors », *Journal of the American Statistical Association*, Juin, 77 (378), pp. 407-415.
- BEDARD, K. 2003. « School Quality and the Distribution of Male earnings in Canada », *Economics of Education Review*, 22(4), pp. 395-407.
- BUCHINSKY, M. 1994. « Changes in the US Wage Structure 1963-1987 : Application of Quantile Regression », *Econometrica*, Mars, 62(2), pp. 405-458.
- BUCHINSKY, M. 1998. « Recent Advances in Quantile Regression Models : A Practical Guideline for Empirical Research », *The Journal of Human Resources*, Winter, 33(1), pp. 88-126.
- CADE, B.S et NOON, B. R. 2003. « À Gentle Introduction to Quantile Regression for Ecologists », *Frontiers in Ecology and Environment*, 1(18), pp. 412-420.
- CIZEK, P. 2003. « Quantile Regression ». *Xplore Application Guide*, édité par W. HARFLE, Z. HLAVKA et S. KLINKE, Springer-Verlag : Heidelberg, pp.19-49.
- DORMONT, B. 1999. « Introduction à l'économétrie », ed. Montchrestien.
- FALARIS, E. M. 2003. « À Quantile Regression Analysis of Wages in Panama ». University of Delaware.
- GARCIA, J., HERNANDEZ, P. J. et LOPEZ-NICOLAS, A. 2001. « How Wide is the Gap ? An Investigation of Gender Wages Differences Using Quantile Regression », *Empirical Economics*, 26, pp. 58-68.
- GREENE, W. H. 2003. « Econometric Analysis », 5ème édition, Prentice Hall International, London.
- HECKMAN, J. J. 1979. « Sample Selection Bias as a Specification Error », *Econometrica*, Janvier, 47(1), pp. 153-161.
- HOGG, R. 1975. « Estimates of Percentile Regression Lines Using Salary Data », *Journal of the American Statistical Association*, Mars, 70(349), pp. 56-59.3
- KOENKER, R. 1995. « Quantile Regression software ». Disponible sur le site <http://www.econ.uiuc.edu/~roger/research/rq/rq.html>
- KOENKER, R. 2005. « Quantile Regression », *Cambridge University Press*.
- KOENKER, R. et BASSETT, G. 1978. « Regression Quantiles », *Econometrica*, Janvier, 46(1), pp. 33-50.
- KOENKER, R. et BASSETT, G. 1982. « An Empirical Quantile Function for Linear Models with iid Errors », *Journal of the American Statistical Association*, Juin, 77(378), pp. 407-415.
- KOENKER, R. et BASSETT, G. 1982-b. « Tests of Linear Hypothese and L1 Estimation », *Econometrica*, novembre, 50(6), pp.1577-1584.
- KOENKER, R. et HALLOCK, K. F. 2001. « Quantile Regression : an Introduction », *Econometrica*, Janvier, 46(1), pp. 33-50.

- KOENKER, R. et MACHADO, A. F. 1999. « Goodness of Fit and Related Inference Processes for Quantile Regression », *Journal of the American Statistical Association*, Décembre, 94(448), pp. 1296-1310.
- KOENKER, R. et XIAO, Z. 2002. « Inference on the Quantile Regression Process », *Econometrica*, Juillet, 70(4), pp. 1583-1612.
- MATA, J. et MACHADO, J. A. F. 1996. « Firm Start-up Size : A Conditionnal Quantile Approach », *European Economic Review*, 40, pp. 1305-1323.
- MELLY, B. 2001. « The Theory and Practice of Quantile Regression », *Universität of St. Gallen*.
- MINCER, J. 1974. « Schooling, Experience and Earnings », New York : National Bureau of Economic research.
- MONTENEGRO, C. 2001. « Wage Distribution in Chile : Does Gender Matter ? À Quantile Regression Approach », The World Bank, *Development Research Group*, Décembre, n°20
- OAXACA, R. 1973. “Male-Female Wage Differentials in Urbal Labour Markets”, *International Economic Review*, n° 14, pp. 693-709.
- POWELL, J. L. 1989. “Estimation of monotonic regression models under quantile restrictions”, in *Nonparametric and Semiparametric Methods in Econometrics*, ed; par W. Barnett, J. Powell et G. Tauchen. *Cambridge University Press*.
- SAAVEDRA, L. A. 2001. « Female Wage Inequality in latin American Labor Markets », The World Bank, *Policy Research Working Papers*, Décembre, n° 2741.
- SCHULZE, N. 2004. « Applied Quantile Regression : Microeconometric, Finance and Environmental Analyses», PhD Dissertation, », *Universität of Tübingen*.