HAL
open science

Vieillissement, dépenses de santé et inégalités de recours aux soins : essais de micro-économétrie appliquée

Hélène Huber

To cite this version:

Hélène Huber. Vieillissement, dépenses de santé et inégalités de recours aux soins : essais de micro-économétrie appliquée. Economies et finances. Université de Nanterre - Paris X, 2006. Français. NNT : . tel-00181627

HAL Id: tel-00181627

https://theses.hal.science/tel-00181627v1

Submitted on 24 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris X Nanterre

UFR Sciences Economiques, Gestion et Informatique

Vieillissement, dépenses de santé et inégalités de recours aux soins : essais de micro-économétrie appliquée

par

Hélène HUBER

Thèse pour le Doctorat en
Sciences Economiques
(Arrêté du 30 mars 1992)

Directrice de Recherche :

Brigitte DORMONT, Professeur à l'Université Paris Dauphine

Jury:

Pierre-Yves GEOFFARD, Directeur de Recherche au CNRS
Alberto HOLLY, Professeur à l'Université de Lausanne, Rapporteur
Myriam KHLAT, Directrice de Recherche à l'INED, Rapporteur
Valérie MIGNON, Professeur à l'Université Paris X
Thomas PIKETTY, Directeur d'Etudes à l'EHESS

Présentée et soutenue publiquement le 4 Décembre 2006

L'Université Paris X Nanterre n'entend donner aucune approbation ou improbation aux opinions émises dans cette thèse: ces opinions doivent être considérées comme propres à leurs auteurs.

Résumé

Vieillissement, dépenses de santé et inégalités de recours aux soins : essais de micro-économétrie appliquée

par

Hélène HUBER

Université Paris X Nanterre,

Brigitte DORMONT, Directrice de Recherche

L'objet de cette thèse est l'analyse des comportements de consommation de soins et de leur impact sur l'évolution et la répartition de la dépense de santé dans la population. Nous menons une analyse sur données individuelles et mettons en œuvre des méthodes originales de micro-simulations afin de mettre en évidence les effets de l'hétérogénéité des comportements au niveau individuel.

Un premier chapitre expose les méthodes économétriques employées lors de l'estimation d'équations explicatives de la consommation de soins sur données individuelles.

Un deuxième chapitre développe une méthode originale visant à comparer les effets du vieillissement démographique à d'autres effets influençant la hausse des dépenses de santé entre les années 1992 et 2000. Nous montrons que la hausse des dépenses est principalement due aux changements de comportement face à la maladie, ces changements de comportements pouvant être en grande partie liés au progrès technique. Le vieillissement démographique n'a qu'un rôle mineur dans l'explication de la hausse des dépenses de santé. De plus, nous montrons que l'état de santé moyen s'améliore.

Un troisième chapitre propose une méthode novatrice de décomposition des inégalités de consommation de soins par facteur. A morbidité donnée, la consommation de soins est concentrée chez les individus les plus riches de la distribution ; or nous montrons que cette inéquité est pour moitié due à l'hétérogénéité des comportements due à la position des individus dans l'échelle des revenus, cette hétérogénéité n'étant pas identifiable dans les décompositions standard. Trois applications sont proposées : la décomposition des inégalités par facteur pour la France en 1998, l'évaluation de l'impact de la CMU-C sur l'évolution des inégalités de consommation de soins, enfin la comparaison des sources d'inégalité entre les personnes âgées (65 ans et plus) et les autres.

Mots-clefs : Dépense de santé - Econométrie - Micro-simulations - Vieillissement - Inégalités

Classification JEL : C15 - C34 - D63 - I11 - I18 - J11

Abstract

Ageing, health expenditures and inequalities in health care use in France : a micro-econometric study

by

Hélène HUBER

University Paris X Nanterre,

Brigitte DORMONT, Supervisor

This work analyzes health care consumption behaviors. We focus on their impact on the changes in health care expenditures and on health care distribution within the French population. We run analyses on individual data and make use of micro-simulations in order to evaluate the effects of the heterogeneity of individual behaviors.

In chapter 1 we outline the econometric methods we use when estimating models of individual health care consumption.

In chapter 2 we describe an original method whose goal is to compare the effects of demographic ageing to other factors driving the growth in health care expenditures that is observed between 1992 and 2000 in France. We show that health expenditures growth is mainly due to changes in behaviors for a given morbidity. These changes are mainly related to technological progress. Ageing plays a minor role in this growth, while health improvements lead to savings.

In chapter 3 we propose an innovative method of the decomposition of health care use inequalities by factors. For a given morbidity, health care consumption is concentrated among the rich ; we show that half of this inequity is due to the heterogeneity of behaviors relative to the rank of individuals in the income distribution. This heterogeneity cannot be identified in the standard decompositions. We propose 3 applications of this method : a decomposition of inequality by factors for France, year 1998, the evaluation of the impact of *CMU* (universal free care) on the changes in health care use inequalities, and a comparison between the sources of inequality between elderly people (65+) and people aged 15 to 64.

Keywords : Health care consumption - Econometrics - Micro-simulations - Ageing - Inequalities

JEL codes : C15 - C34 - D63 - I11 - I18 - J11

A ma famille

Table des matières

Table des figures **vi**

Liste des tableaux **vii**

Introduction **1**

1 Méthodes : spécificités de la micro-économétrie de la santé **13**
1.1 La particularité des dépenses de santé 13
1.2 La transformation *log* 18
1.2.1 Le problème de la retransformation exponentielle 19
1.2.2 Le "homoscedastic smearing estimate" de Duan 23
1.2.3 Estimation directe par les GLM 24
1.3 La prise en compte de la participation 31
1.3.1 Le modèle *Heckit* ou *Tobit généralisé* ou *sample selection model* . . . 32
1.3.2 Le modèle en deux équations indépendantes 37
1.4 Les modèles zero-inflated 39

2 Vieillissement et dépenses de santé : réévaluer l'importance de la démographie **41**
2.1 Introduction 41
2.2 L'apport de la littérature 44
2.2.1 L'impact du vieillissement apparaît marginal au niveau macro-économique 44
2.2.2 La question de la proximité de la mort 46
2.2.3 La proximité de la mort : simple *proxy* de la morbidité ? 48
2.2.4 Le progrès technique 50
2.3 Les données 51
2.3.1 Le système de santé français 51
2.3.2 Les données 52
2.3.3 Analyse descriptive 53
2.4 Lien entre morbidité et âge 59
2.4.1 La sélection d'indicateurs de morbidité exogènes 60
2.4.2 Le lien étroit entre la morbidité et le profil de dépenses par âge . . . 60
2.4.3 Les changements de morbidité à travers le temps 62
2.5 L'approche empirique 64
2.5.1 Spécification économétrique et estimation 64
2.5.2 Prédictions et simulations 68
2.6 Résultats micro-économétriques 72
2.6.1 Estimations 72

2.6.2 Les simulations 73
2.7 Evaluer les effets comparés du vieillissement et des changements de pratiques au niveau macro-économique 79
2.8 Conclusion 81

3 Analyse des inégalités de consommation de soins : l'apport des micro-simulations 83
3.1 Introduction 83
3.2 Décomposition des inégalités de recours aux soins : la méthodologie standard et ses limites 90
3.2.1 L'indice de concentration 90
3.2.2 La décomposition par facteurs de l'indice de concentration 93
3.2.3 Une définition de l'inéquité : le Horizontal Inequity Index 94
3.2.4 Une décomposition nécessairement linéaire 94
3.2.5 L'interprétation des C_k 96
3.2.6 La difficile prise en compte des préférences 98
3.2.7 La limitation à une équation unique expliquant la consommation de soins 98
3.3 Une autre méthodologie par micro-simulations 100
3.3.1 Le choix d'un modèle explicatif du recours aux soins 101
3.3.2 Une décomposition de l'indice de concentration du recours aux soins 102
3.3.3 Estimation des paramètres 103
3.3.4 Simulation des distributions et calcul de leurs indices de concentration 104
3.3.5 Interprétation 108
3.4 Trois applications à l'analyse de l'inégalité de consommation de soins en France 110
3.4.1 Méthode 110
3.4.2 Application 1 : analyse des inégalités de consommation de soins observées en 1998 en France 113
3.4.3 Application 2 : impact de la CMU-C dans la variation d'inégalité observée entre 1998 et 2002 118
3.4.4 Application 3 : comparaison entre les 15-64 ans et les 65 ans et plus, année 1998 122
3.5 Conclusion 125

Conclusion générale 129

A Description des données IRDES 133

B Annexes relatives au chapitre 2 136
B.1 Variables potentiellement utilisées dans les estimations 136
B.1.1 Description sommaire des variables 136
B.1.2 Les groupes d'âge 136
B.1.3 Caractéristiques socio-démographiques 137
B.1.4 Les indicateurs de morbidité (IRDES, notice Appariement année 2000) 138
B.2 La sélection des variables 141
B.2.1 Dépenses de consultations 141
B.2.2 Dépenses de pharmacie 143
B.2.3 Dépenses d'hôpital 144

B.3 Les estimations 146
B.4 Les simulations sur population observée 148

C Annexes relatives au chapitre 3 **149**
C.1 Les données 149
C.2 Calcul du nombre de séances conditionnel 151
C.3 Les variables retenues 152
C.3.1 Variables potentiellement utilisées dans les estimations 152
C.3.2 Variables sélectionnées dans les estimations du chapitre 3 154
C.4 Les distributions simulées 156
C.4.1 Identification des sources d'inégalités de consommation de soins : l'approche par microsimulations 156
C.4.2 Analyse de la robustesse à l'ordre d'introduction des effets 157

Bibliographie **159**

Table des figures

0.1 Dépense de santé moyenne par tranche d'âge, année 1992, données IRDES . 2
0.2 Espérance de vie à la naissance, France, données INSEE 4
0.3 Espérance de vie à 60 ans, France, données INSEE 4

1.1 Distribution des dépenses de pharmacie, année 2002 14
1.2 Distribution des dépenses de pharmacie tronquée à droite, année 2002 . . . 14
1.3 Distribution du *log* des dépenses de pharmacie, année 2002 17
1.4 Graphique de $\lambda(x) = \phi(x)/\Phi(x)$ 35

2.1 Dépense de santé moyenne par tranche d'âge, années 1992 et 2000, données IRDES 42
2.2 Profils de participation et de dépense 56
2.3 Dépenses de pharmacie et d'hôpital (année 2000), tous et "survivants" . . . 58
2.4 Evolution d'indicateurs moyens de morbidité par tranche d'âge 61
2.5 Effet des différents prédicteurs de la dépense pharmaceutique, 2000 62
2.6 Profils simulés 75
2.7 Graphique 2.6f (agrandi) : Dépense pharmaceutique inconditionnelle simulée 76
2.8 Graphique 2.6i (agrandi) : Dépense hospitalière inconditionnelle simulée . . 76

C.1 Consommation observée en 1998 et 2002 par groupe d'âge, consultations . . 149
C.2 Consommation observée en 1998 et 2002 par groupe d'âge, pharmacie . . . 150
C.3 Consommation observée en 1998 et 2002 par groupe d'âge, hôpital 150

Liste des tableaux

0.1 Projections de population en milliers d'individus (scénario central pour la France) INSEE, 2001 1

2.1 Caractéristiques de l'échantillon 54
2.2 Simulations au niveau macro-économique 80

3.1 Nombre moyen de consultations par personne, années 1998 et 2002 110
3.2 Décomposition de l'inégalité observée en 1998 en France 113
3.3 Décomposition de l'inégalité de consommation de séances de généralistes, année 2000, par ECUITY 114
3.4 Nombre moyen de consultations par personne, 2002 119
3.5 Décomposition de l'inégalité observée en 2002 en France 120
3.6 Variation de l'inégalité entre 1998 et 2002 : décomposition par facteur . . . 121
3.7 Nombre moyen de consultations par personne, années 1998 et 2002 122
3.8 Décomposition des inégalités par facteur : comparaison des 15-64 ans et des 65 ans et plus, année 1998 123

B.1 Equation de consommation conditionnelle GLM, dépenses pharmaceutiques : 1992 and 2000 146
B.2 Equation de participation, dépenses pharmaceutiques : 1992 and 2000 . . . 147
B.3 Effectifs de la population française par tranche d'âge, 1992 et 2000 148

C.1 Indices de concentration des distributions simulées - généralistes, 15+ . . . 156
C.2 Apport de chaque étape à l'inégalité - spécialistes, 15+ 158
C.3 Apport de chaque étape à l'inégalité - spécialistes 15+ 158

Remerciements

Ce travail de recherche est l'occasion pour moi de remercier tous ceux qui ont concouru, d'une manière ou d'une autre, à sa réalisation.

En premier lieu, je tiens à remercier le Professeur Brigitte Dormont, qui a dirigé cette recherche. Elle a été pour moi d'une aide inestimable aussi bien sur le plan scientifique qu'humain. Tout au long de ces années de recherches, ses conseils précieux ont encouragé ma réflexion. Pour cela, je lui témoigne toute ma gratitude.

Je remercie également Michel Grignon pour sa collaboration scientifique ainsi que pour son accueil au début de ma thèse au sein de l'IRDES.

Qu'ici soient remerciées les équipes des laboratoires THEMA et EconomiX de l'Université Paris X Nanterre qui m'ont permis de poursuivre sereinement cette recherche en m'apportant des conditions scientifiques et matérielles excellentes.

Je remercie Dominique Polton et Chantal Cases, directrices successives de l'IRDES, qui m'ont ouvert leur locaux, mis à ma disposition leurs données et offert un encadrement précieux. Ainsi, je remercie Paul Dourgnon, Dominique Goldfarb, Catherine Ordonneau, Marc Perronnin, Catherine Sermet, Sandy Tubeuf, ainsi que toute l'équipe pour leur concours et leurs nombreux conseils.

Je remercie tous ceux dont les remarques constructives ont nourri mon travail, lors de séminaires auxquels j'ai pu présenter mes recherches : merci à Jan Askildsen, Pierre-Yves Geoffard, Alberto Holly, Andrew Jones, Myriam Khlat, Audrey Laporte, Claire Lelarge, Michel Lubrano, Owen O'Donnell, Lise Rochaix, Joseph Terza, Eddy Van Doorslaer.

Un travail de thèse nécessite de se réaliser dans une ambiance agréable mais néanmoins studieuse : je remercie ici les doctorants des laboratoires THEMA et EconomiX pour leur bonne humeur et leurs encouragements tout au long de ces années de recherches.

De manière plus personnelle, je remercie ici ma famille qui m'a toujours entourée et encouragée pour me permettre d'aller au bout de mon travail. Leur soutien sans faille et leur disponibilité n'ont eu d'égale que leur affection.

Introduction

La gestion du système de soins français est caractérisée par deux préoccupations principales : d'une part la maîtrise des coûts, d'autre part l'équité de la délivrance des soins. Dans ce cadre, le vieillissement démographique est un enjeu de société car il soulève la question de la pérennité des systèmes de protection sociale. De fait, le vieillissement démographique est souvent rendu responsable de l'augmentation des dépenses de santé observée en France. La question de l'équité du recours aux soins est de même un enjeu de politique publique, dont témoigne la récente instauration de la Couverture Maladie Universelle renforçant le caractère égalitaire du système.

Cette thèse consiste en deux essais de micro-économétrie appliquée visant à analyser d'une part la question du rôle du vieillissement dans l'augmentation des dépenses de santé, d'autre part la question des sources de l'inégalité de consommation de soins observée en France. Les méthodes originales proposées fondées sur des micro-simulations mettent en évidence l'importance cruciale des comportements individuels dans l'évolution et la répartition de la dépense de santé dans la population.

Année	Population totale	Population 65+	% 65+
2000	58 744	9 413	16,02
2010	61 061	10 362	16,97
2020	62 734	13 188	21,02
2030	63 927	15 908	24,88
2040	64 468	18 033	27,97
2050	64 032	18 674	29,16

Tab. 0.1: Projections de population en milliers d'individus (scénario central pour la France) Insee, 2001

L'augmentation du poids des personnes âgées dans la population (voir tableau 0.1) est traditionnellement expliquée par la combinaison de trois phénomènes : la hausse de

l'espérance de vie, la baisse de la fécondité, et l'effet des *baby-boomers* (cohortes nombreuses nées dès après la fin de la Seconde Guerre Mondiale, soit à partir de l'année 1946). La prise en compte des spécificités de ces tranches d'âge constitue un enjeu crucial des politiques publiques, tant en matière de gestion des retraites qu'en matière de gestion du système de santé. Les personnes âgées sont en effet celles qui consomment en moyenne le plus de soins de santé, ce que l'on peut constater sur le graphique suivant, qui présente la dépense moyenne de santé (non liée à la dépendance) par tranche d'âge, calculée pour l'année 1992 sur données françaises (l'Appariement de l'IRDES[1]).

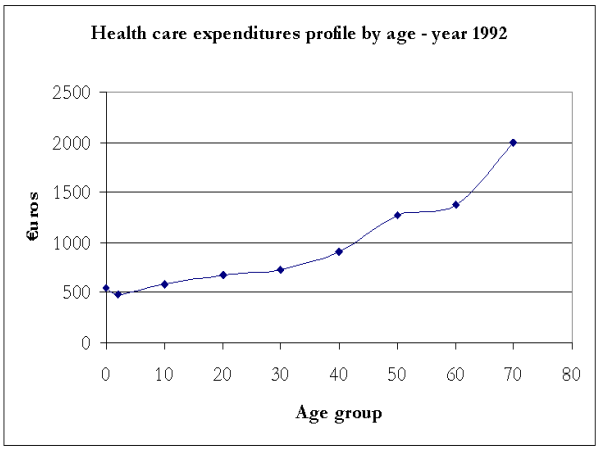


FIG. 0.1: Dépense de santé moyenne par tranche d'âge, année 1992, données IRDES

Le vieillissement : un défi pour le système de protection sociale

Le spectre du vieillissement de la population est souvent invoqué par les médias pour illustrer l'inéluctabilité de la dérive des dépenses de santé que les gouvernements successifs ont cherché à contenir[2]. La Sécurité Sociale, qui comporte quatre branches (famille, vieillesse, accidents du travail et maladie) a été victime en 2004 d'un déficit de 14 milliards d'euros, principalement dû à la branche maladie, dont le déficit s'élevait à plus de 13 milliards d'euros. Les personnes âgées bénéficient en effet de l'Assurance Maladie au même titre que le reste de la population, et il n'existe pas pour elles en France de régime particulier tel *Medicare*[3] aux Etats-Unis. Le poids grandissant des personnes âgées dans

[1] Institut de Recherche et de Documentation en Economie de la Santé, 10 rue Vauvenargues, Paris XVIIIe. `www.irdes.fr`

[2] Une description détaillée des plans mis en œuvre est disponible dans Delattre (2000)

[3] *Medicare* est un système d'assurance santé conçu pour les personnes âgées, sous conditions de ressources, instauré aux Etats-Unis.

la population, associé au fait que ce sont elles qui en moyenne dépensent le plus de soins de santé, pourraient ainsi rendre le vieillissement démographique responsable de l'impossibilité apparente de freiner la hausse des dépenses de santé.

Comment définir le vieillissement ?

Selon le TLF (Trésor de la Langue Française informatisé (CNRS 2006)), la définition démographique du vieillissement, que nous retiendrons ici, est la suivante :

> Vieillissement de la/des population(s). Augmentation du nombre des personnes âgées par rapport à la population totale.

La première définition proposée par le TLF est cependant :

> Fait de vieillir ou d'avoir vieilli, d'avancer en âge, de s'affaiblir en perdant progressivement ses forces physiques ou morales, ses capacités intellectuelles.

Tandis que la définition biologique proposée est :

> Processus par lequel un organisme humain subit une série de transformations entraînant la dégénérescence de certaines cellules, ce qui provoque l'affaiblissement et le ralentissement des fonctions vitales et des modifications d'ordre physique, physiologique et psychique.

La notion de vieillissement est donc connotée extrêmement négativement, ce que l'on retrouve dans la plupart des dictionnaires ainsi que dans l'usage commun qui est fait du terme. Pour A. Sauvy, "un pays sans jeunesse est voué à devenir un pays de vieilles gens ruminant de vieilles idées dans de vieilles maisons" (Sauvy 1953). La définition démographique est toutefois exempte d'une telle connotation.

Le vieillissement de la population (souvent défini comme l'augmentation de la proportion et du nombre des personnes de 65 ans et plus) est ainsi perçu comme une source d'inquiétude tant pour la viabilité du système des retraites que pour celle du système d'Assurance-Maladie. Les personnes âgées sont souvent rendues responsables de la hausse continue des dépenses de santé que les pays développés dans leur ensemble connaissent depuis des années. Et ce, alors même que l'un des indicateurs les plus habituellement retenus pour juger de l'efficacité d'un système de santé est l'espérance de vie de la population concernée.

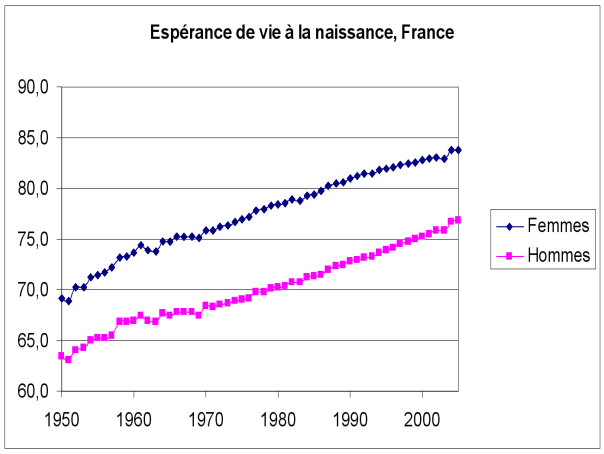


FIG. 0.2: Espérance de vie à la naissance, France, données INSEE

Les indicateurs du vieillissement Si les gains d'espérance de vie à la naissance augmentent chaque année (voir figure 0.2), les gains en espérance de vie dans les pays développés se font désormais sur l'espérance de vie à 60 ans, plus que sur l'espérance de vie à la naissance. En effet, alors qu'auparavant les gains en espérance de vie étaient principalement réalisés grâce à la diminution de la mortalité infantile, les gains en taux de mortalité se font dorénavant aux âges les plus élevés (voir Cutler et Meara (2003) dans le cas des Etats-Unis).

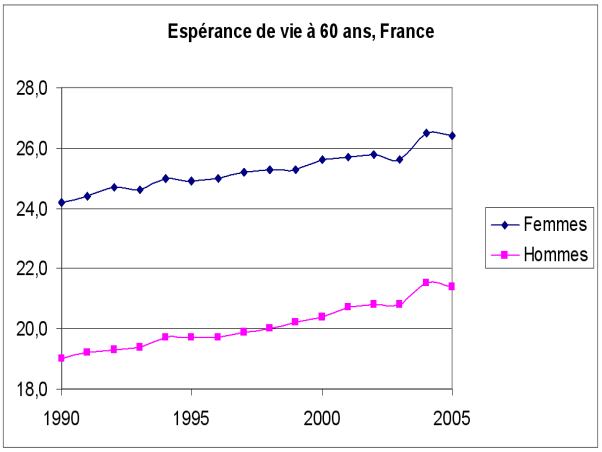


FIG. 0.3: Espérance de vie à 60 ans, France, données INSEE

Or l'espérance de vie est un indicateur du moment, très sensible à la variation des conditions de vie une année donnée (Brouard 2005). En effet, l'espérance de vie d'une population une année donnée N est estimée à partir d'une cohorte fictive qui naîtrait l'année N, et qui serait soumise toute sa vie aux taux de mortalité par âge observés l'année N. Il s'agit donc d'un indicateur très réactif aux variations des taux de mortalité une année donnée.

Un autre indicateur démographique pouvant refléter le vieillissement, la *vie moyenne*, est beaucoup moins réactif : il s'agit de la durée de vie moyenne calculée sur une cohorte réelle, observée jusqu'au décès de tous les individus. Devoir attendre le décès du dernier individu de la cohorte en question rend ce dernier indicateur peu adapté à l'analyse conjoncturelle immédiate. En France, l'espérance de vie est actuellement supérieure à la vie moyenne. Lors de chocs (guerres, périodes de transition), l'espérance de vie peut chuter considérablement (car elle est calculée avec les conditions du moment) alors que la vie moyenne observée peut demeurer relativement stable. L'indicateur espérance de vie ne reflète donc pas une sorte d'âge moyen que les personnes âgées du moment pourraient raisonnablement atteindre. Ainsi, alors que les discours sur le vieillissement se fondent sur l'augmentation de l'espérance de vie, les raisonnements se font en réalité à partir d'indicateurs tout autres, qui ne sont même pas la vie moyenne (non observable un instant donné) mais par exemple l'âge moyen observé dans la population. Or cet âge moyen est fortement influencé par les effectifs de certaines cohortes, comme celle du *baby-boom*, catégorie dont il sera primordial de tenir compte dans les politiques publiques des années à venir. Un autre indicateur plus souvent évoqué est l'effectif de tranches d'âge choisies *a priori*, comme par exemple la tranche d'âge des 65 ans et plus. Or le critère habituel classant les plus de 65 ans dans la catégorie des personnes âgées perd chaque année de sa pertinence : aujourd'hui, une personne de 65 ans peut être en très bonne santé, et refuser de se voir classée dans la catégorie des "personnes âgées", catégorie qui évoque invariablement la dépendance. Le seuil des 65 ans est souvent évoqué comme seuil de la vieillesse, lié au seuil de la retraite professionnelle. Or la simple sortie de la vie active professionnelle ne peut être assimilée à l'entrée dans la dépendance. Il est difficilement soutenable aujourd'hui d'admettre l'équivalence entre vieillissement (au sens d'avancée en âge) et dégradation de l'état de santé.

Pour Hummel (2003), il y a longtemps que le vieillissement au sens de dégradation n'est plus conçu comme un phénomène irrémédiable : la gérontologie contemporaine oppose le vieillissement normal (ou même "réussi"), donc maîtrisé, au vieillissement pathologique (subi).

Le vieillissement constitue-t-il une menace ? Pour le Cacheux et Touzé (2003), "le débat public sur le vieillissement démographique est empreint de pessimisme, voire de catastrophisme". Or ces auteurs, s'intéressant à l'accumulation patrimoniale et à la croissance économique, montrent que le vieillissement peut être facteur de développement, au sens où l'augmentation de la durée de vie anticipée provoque une augmentation de l'activité et de l'épargne, favorable à la croissance et aux générations futures.

Le vieillissement serait ainsi à la fois facteur et conséquence du développement à travers l'augmentation de la durée de vie moyenne. La question de l'augmentation des dépenses de santé liée au vieillissement demeure toutefois un enjeu crucial. Sachant que les besoins et dépenses de santé se concentrent principalement aux âges élevés, le vieillissement constitue-t-il réellement une menace pour le système de santé français ? En d'autres termes, le vieillissement de la population peut-il causer une explosion des dépenses de santé ?

Outre l'augmentation des dépenses de santé, une question cruciale dans la gestion du système de santé est celle de son équité. Le système, conçu comme égalitaire, a pour objectif l'égal accès aux soins. Or une norme d'équité doit être définie afin de juger des performances du système en ce domaine. Seul le recours aux soins est observé, et non l'accès. Les études du groupe Ecuity, qui constituent une référence en ce domaine, retiennent comme norme l'égalité de consommation de soins à morbidité donnée. Une telle définition néglige les inégalités résultant de différences de préférences. Or dans le domaine des soins de santé, la question de laisser ou non les préférences s'exprimer demeure entière. Les questions de santé publique telles la prévention des maladies infectieuses mènent à contrarier les préférences en imposant par exemple la vaccination obligatoire. Outre ces obligations,

le recours aux soins n'est pas encadré. Quel peut alors être le rôle des hétérogénéités de comportement dans l'explication des inégalités de consommation de soins observées ?

*

* *

L'objet de cette thèse est l'analyse des comportements de consommation de soins et de leur impact sur l'évolution et la répartition de la dépense de santé dans la population. Nous mettons en œuvre des méthodes originales de micro-simulations afin de mettre en évidence les effets de l'hétérogénéité des comportements au niveau individuel. Dans le cadre de l'analyse de l'évolution des dépenses de santé, nous comparons l'effet du vieillissement à l'effet des changements de comportement et montrons que ces derniers ont une importance cruciale. L'hétérogénéité des comportements est également un facteur explicatif important de l'inégalité de consommation de soins observée.

L'analyse se fera sur des données d'enquête individuelles françaises issues des bases de données de l'IRDES[4]. La nature des données (enquête sur population générale) nous mène à nous intéresser uniquement aux soins de court terme (*cure*) et non aux dépenses liées à la dépendance (*care*). Notre analyse traitera uniquement des dépenses et non de leur financement, sujet déjà largement traité par ailleurs. Le plan adopté est le suivant :

Le chapitre 1 expose les méthodes économétriques employées pour l'estimation d'équations explicatives de la consommation de soins sur données individuelles. Lors de la modélisation des dépenses de santé, des techniques d'estimation particulières doivent en effet être mises en œuvre afin de tenir compte de la particularité des dépenses de santé, notamment l'asymétrie des distributions et la présence d'un grand nombre d'observations nulles. Les différentes méthodes d'estimation les plus couramment employées dans la littérature sont exposées.

[4]Institut de Recherche et de Documentation en Economie de la Santé, 10 rue Vauvenargues, Paris XVIIIe. `www.irdes.fr`

Le chapitre 2 développe une méthode originale visant à comparer les effets du vieillissement démographique à d'autres effets influençant la hausse des dépenses de santé. La hausse des dépenses de santé peut en effet être due à de nombreux facteurs, tels le vieillissement démographique, les changements de pratiques du côté de l'offre et de la demande, ou les changements de morbidité. Il est difficile de distinguer l'impact respectif de ces facteurs dans l'évolution des dépenses de santé. L'objet de ce chapitre est d'évaluer l'impact respectif des changements de démographie, des changements de morbidité, et des changements de pratiques à morbidité donnée sur la hausse des dépenses de santé observée en France. Nous montrons en effet que la hausse des dépenses de santé provient d'une part de changements dans la proportion et l'effectif de chaque tranche d'âge de la population, d'autre part de changements dans la dépense moyenne encourue par chacune de ces tranches d'âge, autrement dit de changements du profil de dépense par âge au cours du temps. Ce dernier effet est de loin le plus important. La méthode proposée permet de distinguer dans la variation du profil de dépense par âge ce qui est dû aux changements de morbidité de ce qui est dû aux changements de pratiques. On appelle *pratiques à morbidité donnée* la réponse en terme de dépense à un degré de morbidité donné, toutes choses égales par ailleurs.

L'étude vise à décomposer le passage du profil de dépense observé en 1992 au profil de dépense observé en 2000. Les estimations permettent de simuler des profils de dépense intermédiaires sous différentes hypothèses successives, pour les dépenses de consultations, de pharmacie, et d'hôpital. On met en œuvre pour cela une méthode de micro-simulations sur données individuelles.

Le premier profil de dépense intermédiaire est obtenu en simulant ce que les individus observés en 1992 auraient coûté s'ils avaient connu les pratiques de l'année 2000. Un deuxième profil de dépense intermédiaire est ensuite obtenu en simulant ce que les individus observés en 1992 auraient coûté s'ils avaient connu les pratiques de l'année 2000 *et* s'ils avaient connu la morbidité moyenne de leur tranche d'âge qui a été observée en 2000. Nos résultats montrent que l'effet des changements de pratiques à morbidité donnée est un déplacement vers le haut du profil de dépenses, soit une augmentation de la dépense

moyenne pour chaque tranche d'âge considérée. En revanche, l'effet des changements de morbidité est un déplacement vers le bas du profil : les changements de morbidité survenus entre 1992 et 2000 provoquent une baisse de la dépense moyenne par tranche d'âge, à pratiques données.

Les profils simulés sont ensuite appliqués à la structure par âge de la population française, afin de comparer l'effet des changements de profil de dépense à l'effet démographique. Au niveau macro-économique, l'augmentation de la dépense totale due aux changements démographiques est près de 4 fois plus faible que l'effet de changement de profil dû aux changements de pratiques. Les résultats montrent ainsi que le vieillissement n'explique que très partiellement la croissance observée des dépenses de santé. Le fait d'avoir pu estimer un modèle en deux parties (participation puis consommation conditionnelle) permet d'isoler les effets dûs à la participation des effets dûs à la consommation conditionnelle, plutôt liés à l'offre. Les résultats montrent que pour la dépense ambulatoire (consultations et pharmacie), les changements de pratiques à morbidité donnée sont liés à la dépense conditionnelle, et non à la participation : ils ne sont pas liés à des changements de comportement de la part des patients. Les simulations portant sur les dépenses de pharmacie montrent que le déplacement du profil de dépense est principalement dû au côté de l'offre, c'est-à-dire au progrès technique. L'évaluation montre qu'au niveau macro-économique, la réduction de dépense due aux changements de morbidité contrecarre le choc positif dû aux changements dans la structure démographique. Le vieillissement n'est donc pas synonyme de dégradation : le fait que les changements de morbidité induisent une réduction des dépenses suggère que l'état de santé s'améliore à tout âge, et particulièrement chez les personnes âgées. Ce chapitre montre donc que le vieillissement démographique ne constitue qu'une part très faible de l'augmentation des dépenses de santé constatée entre les années 1992 et 2000, et que l'état de santé s'est amélioré entre ces deux années, y compris pour les personnes âgées. Le principal facteur de croissance des dépenses n'est pas le vieillissement démographique ou la dégradation de l'état de santé, mais bien les changements de pratiques à morbidité donnée, autrement dit les changements de comportement face à la maladie. Notre étude a pu montrer que ces changements sont principalement dus au côté de l'offre dans le cas des dépenses pharmaceutiques, suggérant un large effet du

progrès technique.

Le chapitre 3 traite plus précisément de l'équité de la dépense de soins. Le chapitre 2 a en effet montré que la dépense de santé au niveau individuel était déterminée principalement par l'état de santé individuel, ainsi que par les pratiques à morbidité donnée. Ces pratiques sont mises en évidence dans le chapitre 2 sous la forme d'effets moyens par groupe d'âge, calculés sur tout l'échantillon. Rien n'est dit cependant sur l'éventuelle hétérogénéité de ces effets en fonction du revenu des individus. Or une littérature abondante sur les inégalités de santé a montré qu'un facteur explicatif important des inégalités de santé était le statut socio-économique, et en particulier le revenu (plus précisément, le rang de l'individu dans la distribution des revenus). L'état de santé individuel est ainsi potentiellement hétérogène selon la position de l'individu dans la distribution des revenus. De même, une littérature abondante concernant les inégalités de recours aux soins a montré qu'elles pouvaient s'expliquer par des facteurs socio-économiques, et en particulier le revenu. Les pratiques évoquées au chapitre 2 peuvent donc également ne pas être homogènes selon le niveau de revenu de l'individu observé. Ce chapitre portera ainsi sur l'identification des sources de l'inégalité de consommation de soins observée en fonction du revenu, et sa décomposition par facteur.

Nous proposons dans ce chapitre une méthode originale de décomposition des inégalités par facteurs, l'indice d'inégalité choisi étant l'indice de concentration. Les décompositions le plus couramment employées se fondent sur la décomposition de l'indice d'inégalité d'une distribution de consommation de soins en fonction des indices d'inégalités de tous les facteurs explicatifs de cette consommation. Or nous avons montré dans le chapitre 2 l'importance des comportements relatifs aux pratiques à morbidité donnée. L'aspect novateur de notre méthode est la possibilité de distinguer dans l'inégalité observée d'une part l'effet de la distribution selon le revenu des variables explicatives dans l'échantillon (ce qui est présent dans la méthode standard), d'autre part l'effet de l'hétérogénéité des coefficients estimés par quantile de revenu. Ce dernier effet ne peut pas être identifié dans les approches standard. Dans le cadre de notre méthodologie, sa prise en compte permet d'identifier finement l'impact de l'hétérogénéité des comportements selon

le niveau de revenu, que ce soit du côté de la demande ou du côté de l'offre.

Nous nous basons pour ces analyses sur des distributions de consommation de soins, simulées sous différentes hypothèses, desquelles nous calculons l'indice d'inégalité. Nous estimons pour cela un modèle explicatif de la consommation de soins, comprenant des variables socio-économiques pertinentes ainsi que des indicateurs de morbidité exogènes, avant de proposer trois applications.

La première propose une décomposition de l'inégalité de recours aux soins par facteurs observée en France pour l'année 1998, et permet de comparer nos résultats avec ceux de la littérature. Nous montrons l'importance de l'hétérogénéité des comportements dans la décomposition des inégalités ; cette hétérogénéité des comportements peut représenter jusqu'à 50% du degré d'inéquité observé.

Une deuxième application est consacrée au problème de l'évaluation de l'impact de la Couverture Maladie Universelle Complémentaire sur la réduction des inégalités de soins en France. Nos résultats montrent que l'introduction de la CMU-C est à l'origine de la quasi-totalité de la diminution de l'indice d'inéquité de consommation de soins observée entre les années 1998 et 2002.

Enfin, une dernière application est consacrée à la comparaison des inégalités de recours aux soins entre les personnes de 65 ans et plus et les personnes âgées de 15 à 64 ans : pour ces deux sous-groupes de population, les sources d'inégalité sont de nature bien différente. Nous montrons en particulier que pour les séances de généralistes, l'hétérogénéité des comportements liés à la participation est pro-riche chez les 15-64 ans et pro-pauvre chez les 65 ans et plus, ce qui suggère une différence de coût d'opportunité entre ces deux sous-groupes, liée au statut d'activité, ou bien la présence d'un effet de génération.

Chapitre 1

Méthodes : spécificités de la micro-économétrie de la santé

1.1 La particularité des dépenses de santé

Dans cette thèse sont étudiées des données relatives aux dépenses de santé individuelles. Or ces données présentent au niveau de leur distribution des caractéristiques qui les éloignent de la loi normale : outre le fait qu'elles sont toutes positives ou nulles, on observe un grand nombre de zéros (jusqu'à 90% des observations pour les dépenses d'hôpital), la distribution est asymétrique et étalée vers la droite, et les queues de distributions sont épaisses. Le graphique ci-dessous montre la distribution des dépenses de pharmacie observée sur données françaises (l'Appariement de l'IRDES pour l'année 2002 ; unité : Euros courants), comprenant 6000 individus et représentatif de la population française. La distribution a été lissée par la méthode du noyau.

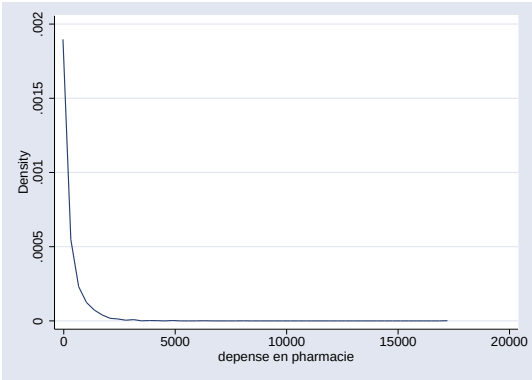


FIG. 1.1: Distribution des dépenses de pharmacie, année 2002

Les dépenses individuelles annuelles de cet échantillon sont comprises entre 0 et 17179 Euros. La moyenne est de 301 Euros, et l'écart-type est de 591. Sur 6000 observations, 859 zéros sont observés, soit une proportion de 14%. Le quantile 95% est situé à la valeur 1295. La représentation graphique de cette distribution tronquée à 1000 permet de mieux la visualiser :

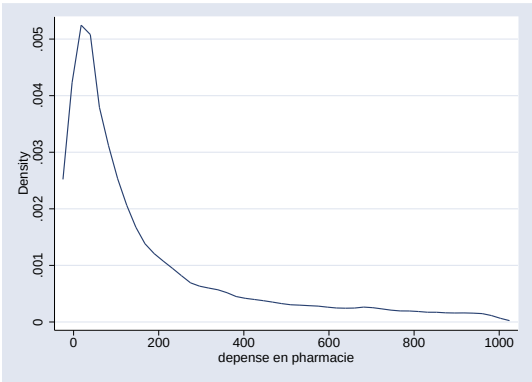


FIG. 1.2: Distribution des dépenses de pharmacie tronquée à droite, année 2002

La *skewness* de cette distribution est de 8,18 et sa kurtosis de 154,52. La *skewness* (notée S) mesure le degré d'asymétrie d'une distribution, alors que la *kurtosis* (notée K mesure son degré "d'écrasement". Pour une distribution de moyenne μ et d'écart-type σ, on a $S = E[(x-\mu)^3]$ et $K = E[(x-\mu)^4]$. On trouve également parfois les S et K normalisés : $S_n = E[(x-\mu)^3]/\sigma^3$ et $K_n = E[(x-\mu)^4]/\sigma^4$. Une distribution symétrique est telle que $S = 0$, alors que $S < 0$ dénote une distribution étalée vers la gauche et $S > 0$ une distribution étalée vers la droite, ce qui est le cas sur nos données. Un K élevé dénotera

une distribution relativement "pointue", alors qu'un K faible dénotera une distribution relativement "écrasée". Pour une distribution normale standard, $S_n = 0$ et $K_n = 3$. Par rapport à une distribution normale standard, la distribution des dépenses de pharmacie observée est donc largement étalée vers la droite, et nettement plus "pointue" (pic dû au grand nombre de zéros).

Ces caractéristiques des dépenses de santé observées au niveau individuel posent problème dès que l'on souhaite évaluer la moyenne des observations. En effet, la distribution des dépenses de santé est asymétrique, avec des outliers[1] extrêmement élevés et la moyenne empirique est très sensible aux valeurs extrêmes. Doit-on éliminer les queues, et si oui, à quel quantile ? Le principe de l'élimination des 5% de valeurs extrêmes, quoique répandu, est largement arbitraire. Les travaux concernant les moyennes robustes apportent des éléments de réponse à ce type de problème lorsque l'on s'intéresse à la moyenne d'une dépense (afin par exemple de rembourser un forfait d'un montant égal à l'espérance des coûts). Par exemple, à propos du problème de l'estimation de la moyenne d'une distribution asymétrique, Marazzi et Ruffieux (1999) proposent le principe de la *moyenne tronquée* : une loi (Gamma, Log Normale ...) est ajustée sur la distribution observée de la variable d'intérêt, puis on choisit un u-quantile de troncature supérieure (par exemple, $u = 0,95$). Le quantile de troncature inférieure est choisi de manière à ce que la moyenne de la loi tronquée coïncide avec la moyenne de la loi non-tronquée. La moyenne de la distribution de la variable d'intérêt limitée aux valeurs comprises entre ces deux bornes fournit une estimation robuste de la moyenne. Cette distribution tronquée de la variable d'intérêt peut permettre l'estimation d'un modèle explicatif de cette variable (Marazzi et Yohai 2004). D'autres méthodes que nous verrons par la suite permettent, en transformant les données, d'éviter la question de l'élimination des queues de distribution.

Du point de vue économétrique, les caractéristiques particulières des dépenses de santé peuvent poser problème dans le cadre de l'estimation d'un modèle explicatif du

[1] Dans un échantillon, on appelle *outliers* les observations anormalement éloignées des autres. Le caractère "anormal" de ces observations est à l'appréciation de l'analyste. Dans le cas des dépenses de santé, les valeurs extrêmement élevées peuvent être qualifiées d'*outliers*.

niveau de dépense individuel. Soit un modèle expliquant le niveau individuel de dépense y par un certain nombre de facteurs rassemblés dans la matrice X. Pour chaque individu i, soit X_i le vecteur ligne contenant les observations des variables explicatives et β le vecteur colonne contenant les paramètres à estimer. Le modèle s'écrit :

$$\forall i \in [1..n], y_i = X_i\beta + u_i \tag{1.1}$$

Sous les hypothèses $E[u] = 0$, $plim\frac{X'u}{n} = 0$ et $plim\frac{X'X}{n} \neq 0$, les estimateurs $\hat{\beta}$ estimés par la méthode des moindres carrés ordinaires (MCO) sont convergents *mais* peu précis. L'inférence sur ces paramètres estimés est également non triviale.

La question de la prise en compte des valeurs extrêmes renvoie à l'hétéroscédasticité des perturbations, souvent présente lors de l'analyse de données en coupe. En l'absence de correction de l'hétéroscédasticité, la méthode des MCO fournit des estimateurs moins précis que l'estimateur des moindres carrés généralisés. On peut donc employer les moindres carrés généralisés (ou quasi-généralisés) mais la difficulté d'identifier la forme de l'hétéroscédasticité rend cet exercice difficile. De plus, l'estimateur des moindres carrés quasi-généralisés est en général biaisé à distance finie[2].

Or, sous certaines hypothèses, dont celle d'hétéroscédasticité, l'estimateur des MCO est sans biais, convergent, et asymptotiquement normal[3]. La normalité est établie grâce au théorème central-limite, même en présence d'hétéroscédasticité[4]. Le fait que les perturbations suivent une loi très différente de la loi normale peut n'être pas gênant. Cependant, en présence d'hétéroscédasticité, l'inférence sur les paramètres estimés est non triviale : en effet, les tests de significativité doivent reposer sur des estimations convergentes de la variance des estimateurs, laquelle dépend des variances des perturbations qui sont difficiles à identifier. En cas d'hétéroscédasticité, on peut utiliser un estimateur convergent de la matrice de variance-covariance asymptotique des estimateurs MCO basé

[2]Voir Dormont (1999), p. 265

[3]Les hypothèses sont $E[u] = 0$, X certaine, X de plein rang, $E[uu'] = \sigma^2\Omega \neq \sigma^2 I_T$, $lim\frac{X'X}{n}$ finie et définie positive, $lim\frac{X'\Omega X}{n}$ finie et définie positive

[4]Voir Dormont (1999), p. 279

sur la matrice de White.

Une dernière caractéristique des dépenses de santé plaidant en la défaveur des *MCO* appliqués tels quels est l'importance des valeurs nulles : la distribution de la variable d'intérêt possède une masse en zéro (les personnes qui ne consomment aucun soin de santé une année donnée) dont il est difficile de rendre compte lors d'une estimation par *MCO*. De plus, le fait que la variable expliquée soit positive ou nulle rend les prédictions des *MCO* inadaptées : rien n'oblige la prédiction du modèle à être positive ou nulle, et une prédiction strictement négative a peu de sens.

La transformation de la variable d'intérêt en *log* est souvent employée pour rapprocher la distribution des dépenses d'une loi approchant la loi normale. Le graphique ci-après montre la densité du *log* de la dépense pharmaceutique (nommée *deppha*) sur le même échantillon que précédemment. On a défini $logdeppha = log(deppha + 1)$: comme $log(1) = 0$, on retrouve après cette transformation la masse d'observations en zéro que l'on observait sur données non transformées.

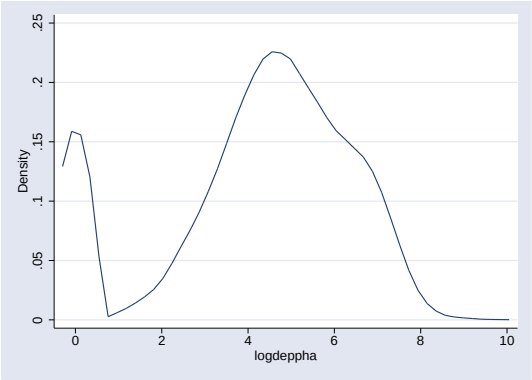


FIG. 1.3: Distribution du *log* des dépenses de pharmacie, année 2002

La distribution *log*-transformée a une distribution apparemment plus proche d'une loi normale, si l'on omet la masse en zéro. Pour la distribution du *log* des dépenses de pharmacie tronquée à zéro, les *skewness* et *kurtosis* sont respectivement $S_n = -0,05$ et $K_n = 2,53$: la transformation *log* permet donc de corriger partiellement l'asymétrie et l'écrasement de la distribution, pour les observations strictement positives. La masse

d'observations nulles peut ainsi faire l'objet d'un traitement séparé. Une pratique très courante, tant en économie de la santé qu'en économie du travail, consiste ainsi à ne pas traiter la variable expliquée y telle quelle, mais à prendre son *logarithme* : ceci est le sujet de la section suivante 1.2. Or la transformation *log* ne résoud pas le problème des nombreux zéros de la distribution des dépenses. La prise en compte de ces nombreux zéros se fait traditionnellement dans le cadre de modèles à deux équations, modélisant la participation d'une part, et la consommation conditionnelle (strictement positive) d'autre part. Citons également dans ce cadre les modèles dits *zero-inflated*, modélisant l'excès de zéros observés dans une distribution. Ces modèles prenant explicitement en compte la masse d'observations en zéro seront l'objet des sections 1.3 et 1.4.

1.2 La transformation *log*

Dans le cadre de l'étude des dépenses de santé observées en France par exemple, la distribution du logarithme de ces dépenses est proche de celle d'une loi normale, ce qui n'est pas du tout le cas de la distribution des dépenses de santé non-transformée. On peut donc effectuer la régression du modèle (1.1) en remplaçant y_i par $Y_i = log(y_i)$. L'emploi d'une telle transformation est habituellement justifié par les raisons suivantes (Manning 1998) : (i) la volonté d'établir l'impact d'une variable explicative de façon multiplicative ou proportionnelle (ii) la volonté d'obtenir un estimateur dont peut se déduire facilement une élasticité (iii) le fait de travailler avec certaines fonctions d'utilité, de demande, de production ou de coût (ex : Cobb-Douglas) (iv) la nécessité de travailler avec des variables dont la distribution est étalée vers la droite ($S > 0$). D'autres transformations de variables à valeurs positives peuvent être utilisées, comme par exemple la transformation inverse, ou la transformation Box-Cox (Box et Cox 1964), où la variable transformée Y se définit par $Y = \dfrac{y^\lambda - 1}{\lambda}$ si $\lambda \neq 0$ et $Y = ln(y)$ sinon. Le λ peut être estimé par maximum de vraisemblance, où la variable transformée Y suit une loi normale. Cependant, malgré leurs propriétés intéressantes, leur usage en économie de la santé reste limité et la transformation *log* est la norme, pour des raisons essentiellement pratiques.

Dans le cas de l'explication des dépenses de santé, l'argument majeur est la réduction de la *skewness* et de la *kurtosis*. Grâce à la réduction de l'hétéroscédasticité, les estimateurs obtenus sont plus précis, permettant d'obtenir les estimations convergentes nécessaires aux tests de significativité des coefficients des variables explicatives. Cependant, les impératifs de communication des résultats ne permettent pas toujours d'annoncer une prédiction des dépenses en *log*-euros. Il est nécessaire de re-transformer les prédictions. Comme nous allons le voir ci-dessous, le problème n'est pas trivial : il ne suffit pas de simplement prendre l'exponentielle de la dépense prédite en *log*.

1.2.1 Le problème de la retransformation exponentielle

Soit un modèle spécifié et estimé sous la forme

$$\forall i \in [1..n], Y_i = X_i\beta + u_i \ avec \ Y_i = log(y_i) \ , \tag{1.2}$$

On peut retrouver la variable expliquée y_i pour tout $i \in [1..n]$ de la façon suivante :

$$y_i = \exp(Y_i) = \exp(X_i\beta + u_i) = \exp(X_i\beta)\exp(u_i) \ . \tag{1.3}$$

Si X_i et u_i sont indépendantes, c'est-à-dire en supposant l'exogénéité des variables explicatives, on obtient :

$$E[y_i] = E[\exp(Y_i)] = E[\exp(X_i\beta + u_i)] = E[\exp(X_i\beta)]E[\exp(u_i)] = E[\exp(X_i\beta)]\Psi_i \tag{1.4}$$

avec $\Psi_i = E[\exp(u_i)]$.

Une fois l'estimation du modèle (1.2) réalisée, on peut formuler une prédiction de la manière suivante :

$$\hat{E}[y_i] = E[\exp(X_i\hat{\beta})]\hat{E}[\exp(u_i)] = \exp(\hat{Y}_i)\hat{\Psi}_i \ , \tag{1.5}$$

où $\hat{\Psi}_i$ est un estimateur de $\Psi_i = E[\exp(u_i)]$. Cette dernière expression montre qu'il ne suffit pas de prendre l'exponentielle de la prévision à l'échelle transformée pour

trouver la prévision à l'échelle non transformée. Il faut tenir compte de l'estimation du facteur de retransformation $\hat{\Psi}_i$. Dans l'hypothèse où les perturbations suivent une loi normale centrée et réduite : $\forall i \in [1..n], u_i \hookrightarrow \mathcal{N}(0, \sigma^2)$, le facteur de retransformation Ψ_i s'écrit :

$$\forall i, \Psi_i = E[\exp(u_i)] = \exp(\frac{\sigma^2}{2}) \tag{1.6}$$

Ce facteur étant indépendant de i, on le notera Ψ.

<u>Démonstration</u>

Soit une variable X suivant une loi log-normale. La variable $Z = ln(X)$ suit une loi normale, d'espérance m et d'écart-type σ. Soit f la densité de X. f est définie pour tous réels a et b strictement positifs tels que $a < b$ par :

$$P(a < x < b) = \int_a^b f(x)dx \ .$$

Or ln est une fonction strictement croissante, donc comme $Z = ln(X)$ suit une loi normale, on a :

$$\begin{aligned}
P(a < x < b) &= P(ln(a) < ln(x) < ln(b)) && (1.7)\\
&= P(ln(a) < z < ln(b)) && (1.8)\\
&= \int_{ln(a)}^{ln(b)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(z-m)^2\right) dz && (1.9)\\
&= \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(ln(x)-m)^2\right) \frac{1}{x} dx && (1.10)
\end{aligned}$$

en effectuant le changement de variable $z = ln(x)$ dans l'expression de l'intégrale. On a donc comme densité pour x, avec x à valeurs strictement positives :

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(ln(x)-m)^2\right) \ .$$

Le calcul de l'espérance d'une variable aléatoire X suivant une loi log-normale est donc le suivant, en effectuant le changement de variable $z = ln(x)$:

$$E(X) = \int_0^{+\infty} x \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(lnx - m)^2\right) dx \tag{1.11}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} \exp\left(\frac{-1}{2\sigma^2}(lnx - m)^2\right) dx \tag{1.12}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(z) \exp\left(\frac{-1}{2\sigma^2}(z - m)^2\right) dz \tag{1.13}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2\sigma^2}(z^2 - 2(m + \sigma^2)z + m^2)\right) dz \tag{1.14}$$

Or on peut écrire :

$$z^2 - 2(m + \sigma^2)z + m^2 = (z - [m + \sigma^2])^2 - [m + \sigma^2]^2 + m^2 \ .$$

On en déduit :

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(-[m + \sigma^2]^2 + m^2)\right) \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2\sigma^2}(z - [m + \sigma^2])^2\right) dz$$

Or $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2\sigma^2}(z - [m + \sigma^2])^2\right) dz$ serait l'intégrale de la densité d'une variable normale d'espérance $m + \sigma^2$ et de variance σ^2, donc égale à 1. L'expression $\int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2\sigma^2}(z - [m + \sigma^2])^2\right) dz$ est donc égale à $\sigma\sqrt{2\pi}$. On a donc :

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{1}{2\sigma^2}(m^2 + \sigma^4 + 2m\sigma^2 - m^2)\right) \sigma\sqrt{2\pi} \tag{1.15}$$

$$= \exp(m + \sigma^2/2) \tag{1.16}$$

Dans le cas où $u_i \hookrightarrow N(0, \sigma^2)$, on a donc $E[\exp u_i] = \exp(\frac{\sigma^2}{2})$.

c.q.f.d.

Pour revenir à notre modèle, un estimateur naturel de Ψ est donc

$$\forall i \in [1..n], \hat{\Psi} = \exp(\frac{\hat{\sigma}^2}{2}) \tag{1.17}$$

où $\hat{\sigma}$ est l'écart-type estimé des perturbations. Si les perturbations suivent des lois normales mais qu'elles sont hétéroscédastiques, on peut chercher à identifier la forme de l'hétéroscédasticité en estimant une fonction de variance exprimant σ en fonction des covariables X : $E[u_i^2] = \sigma_i^2 = \sigma^2(X_i)$. Dans ce cas, plus général, on a :

$$\forall i, E[y_i|X_i] = \exp(X_i\beta)\exp(\frac{1}{2}\sigma^2(X_i)) \tag{1.18}$$

Dans ce cas, l'interprétation des paramètres estimés doit aussi tenir compte de la transformation *log*. Dans le cas de perturbations normales hétéroscédastiques, on a donc, pour toute variable explicative x_j (on omet l'indice individuel i) (Manning 1998) :

$$\frac{\partial E[y|x]}{\partial x_j} = E[y|x](\beta_j + \frac{1}{2}\frac{\partial \sigma^2}{\partial x_j}) \tag{1.19}$$

L'impact de la variable x_j sur y ne se résume donc pas seulement à son coefficient estimé β_j, mais tient compte de l'influence de x_j sur la variance de la perturbation qui, du fait de la retransformation, joue sur l'espérance de y. Remarquons que ce problème se pose car σ_i est exprimé en fonction des variables X_i du modèle : si σ_i était exprimé en fonction d'autres variables W_i, on aurait $\frac{\partial \sigma^2}{\partial x_j} = 0$. L'inférence n'est donc pas immédiate. Les écart-types de ces effets doivent être calculés, car il est possible qu'une variable explicative ait une influence significative sur la variable transformée, alors que cet impact serait non-significatif sur la variable non transformée[5]

Sous l'hypothèse d'homoscédasticité, une interprétation de l'effet partiel $\frac{\partial E[y|x]}{\partial x_j}$ mentionné équation (1.19) est l'effet de la variable x_j exprimé en *pourcentage* : β_j est l'approximation du pourcentage de variation occasionnée à y_i par une variation d'une unité de la variable x_j.

[5] Ai et Norton (2000) proposent une procédure de calcul de la variance des effets partiels des covariables X sur y, au niveau retransformé, dans le cadre hétéroscédastique.

Si la loi des perturbations n'est pas la loi normale, la retransformation faite grâce à l'expression (1.17) donne un résultat biaisé : c'est la raison pour laquelle Duan (1983) a proposé une expression de Ψ_i dans le cas d'une loi quelconque (voir section suivante).

1.2.2 Le "homoscedastic smearing estimate" de Duan

Dans le cas où la loi des perturbations est inconnue mais homoscédastique, Duan (1983) a proposé une estimation non paramétrique de Ψ_i connue sous le nom de *homoscedastic smearing estimate retransformation* :

$$\hat{\Psi} = \frac{\sum_{i=1}^{n} e^{\hat{u_i}}}{n} \ . \tag{1.20}$$

En effet, d'après (1.2), avec $Y = ln(y)$, on a :

$$E[y] = E[\exp(X\beta + u)] = \int \exp(X\beta + u)\, dF(u) \ ,$$

avec F la fonction de répartition de la loi des perturbations. Si cette loi est inconnue, on peut l'estimer par la fonction de répartition empirique $\hat{F}_n$ des résidus $\hat{u_i}$ estimés grâce à l'estimation de (1.2) par les moindres carrés ordinaires :

$$\forall \alpha, \ \hat{F}_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} I\{\hat{u_i} \leq \alpha\}$$

Un estimateur de $E[y]$ est donc :

$$\hat{E}[y] = \int \exp\left(X\hat{\beta} + u\right) d\hat{F}_n(u) \tag{1.21}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \exp\left(X\hat{\beta} + \hat{u_i}\right) \tag{1.22}$$

$$= \exp\left(X\hat{\beta}\right) \frac{1}{n} \sum_{i=1}^{n} \exp(\hat{u_i}) \tag{1.23}$$

$$= \exp\left(X\hat{\beta}\right) \hat{\Psi} \tag{1.24}$$

Duan (1983) montre que sous certaines conditions de régularité (notamment, perturbations iid[6]), l'estimateur *smearing* est convergent.

Cependant, cet estimateur $\hat{\Psi}$ (le *homoscedastic smearing estimate*) n'est comme son nom l'indique applicable qu'uniquement dans le cas où l'homoscédasticité des résidus estimés est reconnue. Dans le cas contraire, la mise en œuvre de cette méthode est cause de biais dans les prédictions (Mullahy 1998).

L'analyse de données transformées en logarithme ne permet pas toujours de résoudre tous les problèmes rencontrés dans le cas des *MCO* sur données non-transformées. En particulier, les résidus estimés peuvent rester hétéroscédastiques. Comme nous l'avons vu dans la section précédente, alors que l'hétéroscédasticité peut être prise en compte dans le cas normal, la présence d'hétéroscédasticité représente un problème empêchant une application de la méthode de Duan (1983) dans le cas d'une loi inconnue. Il n'existe pas à notre connaissance dans la littérature de solution unanimement acceptée pour résoudre le problème de la retransformation dans ce cas. Empiriquement, la plupart des solutions rencontrées sont des retransformations par sous-groupe : on identifie une variable susceptible d'expliquer l'hétéroscédasticité, par exemple la tranche d'âge, et on calcule un $\hat{\Psi}$ par sous-groupe, en employant la méthode de Duan (expression (1.20)) si la loi est inconnue et l'expression (1.6) si l'on prouve la normalité des perturbations (Manning, Newhouse, Duan, Keeler, et Leibowitz 1987). Dans la célèbre étude de la RAND[7], étudiant l'aléa moral lié au niveau de couverture assurantielle, les sous-groupes étaient identifiés par la souscription à telle ou telle police d'assurance.

1.2.3 Estimation directe par les GLM

L'emploi des *GLM* (*Generalized Linear Models*) permet d'estimer un modèle comparable dans son essence à celui évoqué plus haut, mais en ne se bornant pas à une relation linéaire entre l'espérance de la variable réponse et les variables explicatives. Les *GLM* sont une classe très générale de modèles, dont le modèle linéaire classique avec

[6]identiquement et indépendamment distribuées

[7]Rand Corporation, à l'origine d'une expérience sur les comportements de dépense de soins en présence de différents taux de couverture, appelée *Rand Health Insurance Experiment*, menée entre les années 1974 et 1977 aux Etats-Unis.

perturbations normales (estimé par *MCO* ou par maximum de vraisemblance) n'est qu'un sous-cas. En toute généralité, un *GLM* expliquant la variable y en fonction des variables explicatives X est caractérisé par la formulation suivante : $E[y|X] = \mu$ avec y suivant une loi de la famille exponentielle de moyenne μ, avec $g(\mu) = X\beta$, g étant une fonction de lien inversible (Dobson 1993). Un cas fréquent est la fonction de lien *log* : on a dans ce cas $E[y|X] = \exp(X\beta)$. Une formulation générale de la densité d'une loi de la famille exponentielle est la suivante, la loi étant de paramètre θ :

$$f(y,\theta) = \exp\left(a(y)b(\theta) + c(\theta) + d(y)\right) \tag{1.25}$$

cette forme étant canonique si $a(y) = y$. Par exemple, dans le cas de la loi de Poisson, en notant $f(y,\theta)$ sa densité, on a :

$$\begin{aligned} f(y,\theta) &= \frac{\theta^y e^{-\theta}}{y!} && (1.26)\\ &= \exp\left(ln\frac{\theta^y e^{-\theta}}{y!}\right) && (1.27)\\ &= \exp\left(yln\theta - \theta - ln(y!)\right) && (1.28) \end{aligned}$$

On a donc dans le cas de la loi de Poisson : $a(y) = y$, $b(\theta) = ln(\theta)$, $c(\theta) = -\theta$ et $d(y) = -ln(y!)$.

Une autre écriture de la densité d'une loi de la famille exponentielle sous la forme canonique peut être la suivante[8] :

$$f(y,m) = \exp(A(m) + B(y) + C(m)y) \tag{1.29}$$

où m est la moyenne de la loi. Dans le cas de la loi de Poisson, θ et m coïncident. Le cas du modèle linéaire classique avec l'hypothèse de perturbations normales est la configuration où g est l'identité et y suit la loi normale. On voit que dans les *GLM*, contrairement au modèle linéaire classique :

[8] Elle est utilisée par Gourieroux et Monfort (1995)

1. la variable expliquée y peut suivre une loi quelconque de la famille exponentielle,
2. le lien entre la variable expliquée y et les variables explicatives X ne se limite pas au cas linéaire.

Dans le cas qui nous intéresse, qui est celui de l'explication des dépenses de santé, la fonction de lien est très souvent le logarithme (que nous emploierons ici) : $ln\,(E(y)) = X\beta$. En effet, la variable à expliquer étant une dépense de santé, elle doit nécessairement être positive ou nulle : la spécification $E[y|X] = \exp(X\beta)$ garantit cela. Lors de l'estimation par (pseudo-) maximum de vraisemblance, la loi que suit y est choisie parmi la famille des lois exponentielles. Une telle loi peut être discrète (Binomiale, Poisson, Binomiale Négative) ou bien continue (Normale, Gamma, Normale Inverse).

Par exemple, soit une variable Y suivant une loi de Poisson. La densité f de la variable Y est dans ce cadre définie par :

$$f(y, \lambda) = P(Y = y) = e^{-\lambda}\frac{\lambda^y}{y!} \ ,$$

et l'on a $E[y] = V[y] = \lambda$. Un modèle *GLM* expliquant la variable Y par les variables X avec pour fonction de lien le *log* est spécifié pour tout individu i de la manière suivante :

$$E[y_i] = \exp\left(X_i\beta\right) \ .$$

Remarquons qu'un tel modèle admet que $V(y_i)$ dépend de l'observation considérée : $V(y_i) = E[y_i] = \exp\left(X_i\beta\right)$.

Si les observations i sont indépendantes, la vraisemblance du modèle s'écrit donc de la façon suivante :

$$L = \prod_{i=1}^{n} f(y_i, \lambda) \tag{1.30}$$

$$= \prod_{i=1}^{n} f(y_i, X_i\beta) \tag{1.31}$$

$$= \prod_{i=1}^{n} \exp\left(-e^{X_i\beta}\right) \frac{(\exp(X_i\beta))^{y_i}}{y_i!} \tag{1.32}$$

$$\tag{1.33}$$

Et la log-vraisemblance s'écrit :

$$LL = \sum_{i=1}^{n} -e^{X_i\beta} + y_i X_i\beta - ln(y_i!) \ ,$$

que l'on maximise par rapport à β.

L'intérêt majeur de cette approche est que la prédiction de la dépense se fait naturellement. En effet, $ln\,(E(y)) = X\beta$ donc $E(y) = \exp(X\beta)$, ce qui s'estime par $\hat{E}(y) = \exp(X\hat{\beta})$. On voit qu'ici la retransformation est libérée du facteur multiplicatif Ψ qui apparaissait lorsque l'on estimait par les *MCO* la fonction de dépense en logarithme avec le modèle linéaire classique.

Le choix de la fonction de lien et de la forme de la distribution de la variable réponse est essentiellement motivé par l'observation des données, et certains tests menant à la spécification la plus pertinente. En économie de la santé, la fonction de lien est traditionnellement le *log* (d'autres transformations sont toutefois disponibles, comme la fonction inverse $x \longmapsto \frac{1}{x}$ par exemple). Une fois les données transformées, des tests permettent de déterminer la loi la plus appropriée : il s'agit d'un test de Park décrit par Manning et Mullahy (2001) à partir de Park (1966), précisé plus bas.

La vraie loi que suit la variable réponse y étant inconnue, l'estimation du modèle se fait par pseudo-maximum de vraisemblance (qui revient au maximum de vraisemblance

dans le cas où la loi choisie correspond à la vraie loi). Gourieroux et Monfort (1995)[9] montrent que sous certaines hypothèses, le maximum de vraisemblance avec erreur sur la loi peut fournir une estimation convergente du paramètre d'intérêt, si les pseudo-densités sont de la forme exponentielle linéaire : $f(y, m) = \exp A(m) + B(y) + C(m)y$ où m est la moyenne de la loi.

L'idée générale est que même si la fonction de *log*-vraisemblance est mal spécifiée, du moment que les équations du score sont d'espérance nulle dans le cadre du vrai processus générateur de données, les paramètres estimés seront convergents et asymptotiquement normaux[10]. On voit donc que même si la "vraie" loi n'appartient pas à la famille exponentielle, estimer le modèle grâce au pseudo-maximum de vraisemblance employant une pseudo-densité de la famille exponentielle mènera à un estimateur convergent, sous les conditions standard de régularité sous lesquelles les M-estimateurs sont convergents[11]. Gourieroux et Monfort (1995) montrent également que comme la pseudo-log-vraisemblance à maximiser est du type $\sum A(g(x_i, b)) + B(y_i) + C(g(x_i, b)y$, comme $B(y_i)$ ne dépend pas du paramètre par rapport auquel on optimise, la procédure revient à maximiser $\sum A(g(x_i, b)) + C(g(x_i, b)y$: il n'est pas nécessaire que les observations y_i satisfassent les contraintes liées au support des lois de la pseudo-famille (Gourieroux et Monfort 1995). Par exemple, il est possible de choisir une distribution de Poisson (modélisant originellement une variable de comptage, donc des valeurs entières) pour modéliser des dépenses qui par nature sont continues. En revanche, il est nécessaire de tenir compte des contraintes sur la moyenne : dans le cadre de l'analyse des dépenses de santé, la moyenne conditionnelle $g(x, b)$ doit être à valeurs positives.

Au sein de la famille des lois exponentielles, le choix d'une loi n'est donc pas dicté par la nature de son support, mais plus souvent par l'analyse des caractéristiques des données, et l'identification du lien existant entre la moyenne et la variance de la variable d'intérêt. Suivant l'expression (1.25), l'espérance et la variance de $a(Y)$ (forme générale)

[9] Chapitre sur les M-estimateurs (Tome 1, p. 225), dénommés ainsi car il s'obtiennent en minimisant ou maximisant une fonction objectif qui n'est pas forcément la bonne vraisemblance.

[10] Voir Manning et Mullahy (2001) citant Gourieroux, Monfort, et Trognon (1984).

[11] Voir Gourieroux et Monfort (1995), p. 225. Les conditions de convergence du pseudo-maximum de vraisemblance sont (*i*) les individus dont on tire les observations sont iid (*ii*) l'espace des paramètres est compact (*iii*) les fonctions de densités retenues sont continues et intégrables par rapport à la vraie loi.

peuvent se calculer (Dobson 1993) et sont du type : $E(a(Y)) = \frac{-c'(\theta)}{b'(\theta)}$ et $V(a(Y)) = \frac{b''(\theta)c'(\theta)-c''(\theta)b'(\theta)}{(b'(\theta))^3}$. Ces expressions sont modifiées dans le cas d'une loi canonique : en particulier, pour une variable Y suivant une loi de Poisson de densité $f(y,\lambda) = \frac{\lambda^y \exp -\lambda}{y!}$, on a $E(Y) = V(Y) = \lambda$. L'espérance et la variance sont alors exactement proportionnelles, contrairement à ce que l'on trouve dans le cas d'une loi normale par exemple, où la variance n'a pas de lien avec l'espérance. Dans le cas d'une variable X suivant la loi gamma de paramètres α, β et γ, de densité $f(x) = \frac{(x-\gamma)^{\alpha-1}\exp(-(x-\gamma)/\beta)}{\beta^\alpha\Gamma(\alpha)}$ avec $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-t}dt$, α et β positifs et $x > \gamma$ (avec $1/\beta = \theta$ le paramètre d'échelle et α le paramètre de "forme"), la densité se réécrit dans le cas standard où $\gamma = 0$ de la façon suivante : $f(x) = \frac{x^{\alpha-1}\exp(-x/\beta)}{\beta^\alpha\Gamma(\alpha)}$. L'espérance et la variance s'écrivent $E(X) = \alpha\beta$ et $V(X) = \alpha\beta^2$: la variance est donc proportionnelle au carré de l'espérance dans un rapport α, car on a $V(X) = \frac{1}{\alpha}E(X)^2$. Le test de Park utilise ces caractéristiques pour poser un diagnostic sur les données et choisir la loi la plus proche de ce que l'on observe.

Il est également possible d'estimer les modèles de type *GLM* grâce à la méthode du quasi-maximum de vraisemblance, qui se différencie du pseudo-maximum de vraisemblance (quoique les deux appellations coexistent[12]) de la manière suivante : la distribution n'est pas définie, et seule la fonction d'espérance est spécifiée, la fonction de variance étant définie à partir de la relation qu'elle entretient avec l'espérance. En effet, pour une variable Y_i suivant une loi de la famille exponentielle de moyenne μ_i, on a : $V(y_i) = \phi V(\mu_i)$ avec ϕ le paramètre de dispersion et $V(\mu_i)$ la fonction de variance, qui peut par exemple être du type $V(\mu_i) = 1$, $V(\mu_i) = \mu_i$ ou $V(\mu_i) = \mu_i^2$. La variance peut être définie comme fixe (cas de la loi normale), proportionnelle à l'espérance (cas de la loi de Poisson), proportionnelle au carré de l'espérance (cas de la loi Gamma), ou bien avec un lien quelconque. On peut rapprocher chacun de ces différents cas d'une loi, mais la spécification de ce lien entre espérance et variance n'implique pas de spécification d'une distribution particulière. En effet, spécifier que la variance est proportionnelle au carré de l'espérance (par exemple) n'implique pas que la distribution choisie soit une loi Gamma. Ceci est justifié par le fait que dans le quasi maximum de vraisemblance, on ne travaille pas à partir de l'expression de

[12]La différence conceptuelle entre le quasi- et le pseudo-maximum de vraisemblance est développée par Gourieroux et Monfort (1993)

la vraisemblance qui est inconnue, mais à partir d'une quasi-vraisemblance[13], c'est-à-dire une sorte de score qui possède de bonnes propriétés. La distribution n'est pas imposée, et la spécification est plus souple.

De manière générale, la plupart des travaux de recherche visant à expliquer les dépenses de santé spécifient explicitement la loi adoptée. La procédure est donc une estimation par pseudo-maximum de vraisemblance, qui comme on vient de le mentionner possède de bonnes propriétés. C'est le choix que nous adopterons dans la suite. En observant la relation entre l'espérance et la variance de la variable réponse, on peut restreindre le champ initial qui était la famille des lois exponentielles pour ne retenir qu'un type de loi (Poisson, ou Gamma par exemple). Dans le cas des dépenses de santé, les lois habituellement choisies sont la loi Gamma ou la loi de Poisson, selon les caractéristiques observées des données (on a vu en effet que l'on pouvait employer la loi de Poisson pour des variables de nature continue). L'identification de la relation entre espérance et variance peut se faire grâce au test de Park (Park 1966). Le principe de ce test est le suivant :

1. Régresser la variable réponse y sur les variables explicatives X par MCO
2. Calculer la prédiction $\hat{y}$ et les résidus estimés $\hat{u}$
3. Régresser $ln(\hat{u}^2)$ sur $ln(\hat{y})$ et une constante

En effet, si les paramètres cherchés sont a et b tels que $V(y) = aE(y)^b$, la linéarisation de cette expression donne bien $ln(V(y)) = ln(a) + bln(E(y))$, paramètres qui s'estiment naturellement par la régression expliquée plus haut, et qui sont convergents si la variance est bien de cette forme. Un paramètre estimé $\hat{b} = 1$ signifierait que la variance est exactement proportionnelle à l'espérance, ce qui est caractéristique d'une loi de Poisson. De même, un paramètre estimé $\hat{b} = 2$ signifierait que la variance est proportionnelle au carré de l'espérance, ce qui est caractéristique d'une loi Gamma. Le cas où ce paramètre estimé est significativement différent de ces deux valeurs peut poser problème[14], mais si l'on souhaite conserver la méthode du pseudo-maximum de vraisemblance[15], on peut se

[13](Wedderburn 1974, Nelder et Wedderburn 1972, McCullagh et Nelder 1989, Blough, Madden, et Hornbrook 1999)

[14]Là réside justement l'intérêt des méthodes de quasi-maximum de vraisemblance, permettant un lien *ad hoc* entre espérance et variance

[15]Pseudo-maximum de vraisemblance d'ordre 1, car seule l'espérance est spécifiée, et la variance n'est définie qu'à partir de l'espérance.

livrer à un arbitrage entre loi de Poisson et loi Gamma sur la base de la valeur trouvée pour $\hat{\beta}$.

Dans le cadre de l'analyse des dépenses individuelles de santé, le consensus actuel consiste à préconiser l'utilisation des *GLM* plutôt que les *MCO* sur données transformées. Selon Basu, Manning, et Mullahy (2004), les *GLM* sont la spécification la plus robuste aux variations du processus générateur de données. Le problème de la transformation *log* nécessitant la retransformation n'a été soulevé qu'en économie de la santé, grâce aux expériences de la RAND. En économie du travail, des prédictions sont établies grâce à des estimations où la variable expliquée est spécifiée en *log*, sans autre commentaire sur la façon de procéder.

1.3 La prise en compte de la participation

Les observations montrent que sur une population entière, beaucoup d'individus ont une consommation nulle de services de santé : on n'observe la consommation que de ceux qui consomment effectivement. On peut donc penser à un modèle expliquant les comportements de consommation en deux équations :

1. Modélisation de la décision de consommer des soins (participation),
2. Si la décision est positive, modélisation du niveau de consommation atteint.

Un tel modèle en deux équations peut être mis en œuvre soit par le modèle dit *Heckit* ou *Tobit généralisé*, ou *Sample Selection Model* (SSM), soit par le modèle dit *Two-part model* (TPM). La question du choix entre ces deux types de modèles a donné lieu à une controverse scientifique riche en événements (Dow et Norton 2002, Manning et Mullahy 2001) et qui avait conclu à une supériorité du *TPM*, sur des critères qu'il est nécessaire d'analyser. Leung et Yu (1996) ont quant à eux fourni une règle de choix que nous exposerons dans ce paragraphe.

1.3.1 Le modèle *Heckit* ou *Tobit généralisé* ou *sample selection model*

Ce modèle tient compte d'une éventuelle corrélation entre les perturbations des deux équations. En ce sens, il s'agit d'un modèle tout à fait général. Cependant, les hypothèses portant sur la loi suivie par les perturbations sont finalement assez restrictives.

Enoncé du modèle - justification

Pour chaque individu $i \in [1..n]$, on note P_i sa probabilité de participer au système de soins (expliquée par les variables X_1) et Y_i sa dépense conditionnelle au fait de participer (expliquée par les variables X_2). Ici nous ne tiendrons pas compte de l'éventuel problème de la retransformation, mais il convient de garder à l'esprit que cette composante supplémentaire doit se rajouter au calcul d'une prédiction au cas où la variable Y serait transformée. Pour chaque individu i :

$$1. \quad P_i = \mathbf{I}_{P_i^*>0} \quad avec \quad P_i^* = X'_{1,i}b + u_{1,i} \tag{1.34}$$

$$2. \quad Y_i = \mathbf{I}_{P_i=1}.\,[Y_i^*] \quad avec \quad Y_i^* = X'_{2,i}\beta + u_{2,i} \tag{1.35}$$

où

$$(u_{1,i} \;\; u_{2,i}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right).$$

Remarques Ce modèle représente une variable Y tronquée (Greene 2002, Heckman 1979), ce phénomène de troncature étant directement dû à la réalisation de la variable P de première étape. La prise en compte de la corrélation entre les perturbations des deux équations permet d'analyser le lien entre les deux décisions : si par exemple $\rho < 0$, cela voudrait dire que les individus les moins enclins à consulter sont ceux qui consomment le plus une fois prise la décision de participation. Une probabilité moindre de recours aux soins à morbidité donnée peut ainsi provoquer une consommation plus élevée une fois le recours décidé, à cause de la dégradation d'état de santé provoquée par le recours tardif. Ceci peut être vérifié en particulier chez les personnes à faible revenu (voir chapitre 3).

Estimation

Ce modèle en deux équations peut s'estimer en une seule fois par maximum de vraisemblance, mais on peut aussi adopter une estimation en deux étapes (Heckman 1979). Il faut rappeler toutefois que la seconde équation ne peut s'estimer isolément sur les individus pour qui $y > 0$ à cause de la corrélation entre les perturbations des deux équations. En effet, la variable Y suit une loi normale tronquée, et cette troncature dépend directement de la réalisation de la variable P de première étape. Par exemple, soit f la densité jointe de y et z, on a $f(y, z|z > a) = \frac{f(y,z)}{Prob(z>a)}$: la réalisation de la variable provoquant la troncature a une influence sur la distribution de la variable d'intérêt. Soient deux variables y et z suivant une loi normale bivariée de moyennes μ_y et μ_z, d'écart-types σ_y et σ_z et de corrélation ρ, on a :

$$E[y|z > a] = \mu_y + \rho\sigma_y\Lambda(\alpha_z) \tag{1.36}$$

et

$$V[y|z > a] = \sigma_y^2[1 - \rho^2\delta(\alpha_z)] \tag{1.37}$$

avec $\alpha_z = \frac{a-\mu_z}{\sigma_z}$, $\Lambda(\alpha_z) = \frac{\phi(\alpha_z)}{1-\Phi(\alpha_z)}$, $\delta(\alpha_z) = \Lambda(\alpha_z)[\Lambda(\alpha_z) - \alpha_z]$, et Φ et ϕ représentant respectivement la fonction de répartition et la densité de la loi normale centrée et réduite (Greene 2002).

Dans le cadre du modèle résumé par (1.34) et (1.35), on a donc d'après (1.36),

$$E[Y|P > 0, X_1, X_2] = X_2'\beta + \rho\sigma\frac{\phi(X_1'b)}{\Phi(X_1'b)} \ . \tag{1.38}$$

L'expression $\frac{\phi(X_1'b)}{\Phi(X_1'b)}$ est l'inverse du ratio de Mill, que l'on peut noter λ. Soulignons que dans l'estimation, b et σ_1 ne sont pas identifiables individuellement car dans la vraisemblance seul le rapport $\frac{b}{\sigma_1}$ est présent. Il est d'usage de normaliser la variance de la perturbation de l'équation 1 du modèle : $\sigma_1 = 1$, ce qui donne l'expression (1.38) en notant $\sigma = \sigma_2$. L'expression $\rho\sigma\frac{\phi(X_1'b)}{\Phi(X_1'b)}$ représente le biais de sélection. En effet, dans le cas où il existe une corrélation entre les perturbations des deux équations ($\rho \neq 0$), la réalisation de l'équation (1.34) influe sur la réalisation de l'équation (1.35) et ne pas en tenir compte

biaise l'estimation.

L'estimation en deux étapes a lieu de la manière suivante : (i) estimation de l'équation (1.34) par maximum de vraisemblance (Probit) sur tout l'échantillon (ii) calcul de $\hat{\lambda} = \frac{\phi(X_1'\hat{b})}{\Phi(X_1'\hat{b})}$ (estimateur de λ) (iii) estimation de l'équation

$$Y = X_2'\beta + \alpha\lambda + v \tag{1.39}$$

par *MCO*, sur les individus pour qui $Y > 0$. On remarque ici que le paramètre à estimer α est tel que $\alpha = \rho\sigma$: son signe indique le sens de la corrélation entre les perturbations des deux équations. Comme λ est inconnu, on le remplace par $\hat{\lambda}$, et l'équation réellement estimée est la suivante :

$$Y = X_2'\beta + \alpha\hat{\lambda} + w. \tag{1.40}$$

Le test de la significativité du paramètre α dans la régression (1.40) permet de tester la significativité de ρ, donc de tester la corrélation entre les résidus des deux équations (1.34) et (1.35) du modèle initial.

Remarques

Hétéroscédasticité Les perturbations de l'équation de deuxième étape (1.39) sont hétéroscédastiques. L'événement $[P_i = 1]$ est la réalisation de l'événement $[u_1 > -X_1'b]$. On a donc : $Var(v|P_i = 1) = Var(v|u_1 > -X_1'b)$. v et u_1 étant corrélées, d'après l'expression (1.37), l'expression de $V(v|u_1 > -X_1'b)$ est une fonction de X_1 : les perturbations v sont donc hétéroscédastiques. Le fait d'introduire $\hat{\lambda}$ au lieu de λ introduit également de l'hétéroscédasticité, car dans la perturbation w de (1.40) est introduite la différence $\lambda - \hat{\lambda}$, donc une fonction des variables de première étape. L'estimation est donc convergente, mais pas efficace. Outre cet inconvénient, la mise en œuvre du *homoscedastic smearing estimate* de Duan en cas de variables transformées est problématique à cause de cette hétéroscédasticité.

Non-exogénéité Une autre conséquence du remplacement de λ par son estimation $\hat{\lambda} = \frac{\phi(X_1'\hat{b})}{\Phi(X_1'\hat{b})}$ dans (1.40) est un problème de non-exogénéité. Dans la perturbation w de l'équation (1.40) se trouve non seulement la perturbation v de l'équation (1.39), mais également la différence $\alpha(\lambda - \hat{\lambda})$, corrélée aux variables de seconde étape si X_1 et X_2 sont corrélées. L'estimateur de (1.40) est donc biaisé à distance finie, du fait de la corrélation entre la perturbation w et les variables explicatives de la seconde équation, mais reste convergent.

Problèmes d'identification Une trop forte corrélation entre $\hat{\lambda}$ et les variables de seconde étape X_2 pose des problèmes d'identification (Leung et Yu 1996) et l'interprétation de son coefficient estimé $\hat{\alpha}$ dans (1.40) est difficile. De petites variations dans les données peuvent provoquer des changements importants sur les paramètres : cela mène à des problèmes de robustesse et de manque de précision dans l'estimation. Un graphique de $\lambda(x)$ en fonction de $x \in R$ (cf. graphique 1.4) montre que λ a une forme quasi linéaire : une trop grande proximité entre les variables de première et de seconde étape peut mener à une colinéarité entre $\hat{\lambda}$ et les variables de seconde étape.

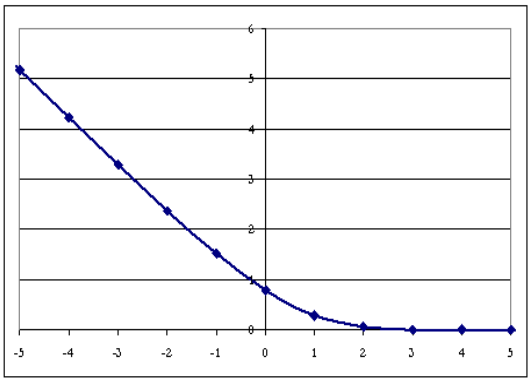


FIG. 1.4: Graphique de $\lambda(x) = \phi(x)/\Phi(x)$

Dans ce cas, il est difficile de juger de la pertinence de l'introduction de $\hat{\lambda}$ dans l'équation de deuxième étape et donc de la pertinence du modèle *sample selection* (Dobson 1993). En revanche, lorsque $\hat{\lambda}$ n'est pas trop corrélé aux variables X_2, le coefficient $\hat{\alpha}$ estimé est un estimateur convergent de $\rho\sigma$. Tester sa significativité permet d'examiner si les perturbations de la première et de la seconde équation sont corrélées.

Pour étudier la corrélation entre $\hat{\lambda}$ et X_2, on calcule le R^2 de la régression par moindres carrés de $\hat{\lambda}$ sur les variables X_2 : la racine carrée de ce R^2 indique le degré de corrélation entre $\hat{\lambda}$ et les variables X_2. La valeur seuil reste à l'appréciation de l'analyste, mais une valeur supérieure à $0,80$ peut indiquer une trop forte corrélation.

Leung et Yu (1996) ont montré grâce à des expériences de Monte-Carlo que la pertinence du *sample selection model* est étroitement liée au degré de corrélation entre le $\hat{\lambda}$ et les variables de seconde étape. Dans le cas où les données ont été générées par un *sample selection model*, une estimation par *two-part model* est source de biais. Cependant, un arbitrage doit être fait entre le biais causé par le *TPM* et la mauvaise précision due à l'emploi du *SSM*. Un critère de choix est l'erreur quadratique moyenne, permettant de comparer des estimateurs entre eux. Soit un paramètre θ dont un estimateur est T, l'erreur quadratique moyenne de cet estimateur est $EQM(T) = E[(T-\theta)^2] = V(T)+(E[T]-\theta)^2 = V(T) + (B(T))^2$, en notant $B(T)$ le biais de l'estimateur T.

Dans le cas d'une forte corrélation entre le $\hat{\lambda}$ et les variables de seconde étape, alors même que les données ont été générées grâce à un *sample selection model*, l'erreur quadratique moyenne est plus faible en estimant un modèle avec deux parties indépendantes qu'en estimant un *sample selection model*. En effet, l'estimateur *SSM* est peu précis à cause de la colinéarité du $\hat{\lambda}$: l'estimateur *TPM*, même biaisé, a une erreur quadratique moins élevée. En revanche, dans le cas d'une corrélation faible, et dans les mêmes conditions, le *sample selection model* est approprié. Cet article constitue une réponse à Manning, Duan, et Rogers (1987), qui avaient pensé prouver la supériorité des modèles en deux parties indépendantes, quel que soit le modèle avec lequel avaient été générées les données. Ils concluaient toujours à la supériorité du modèle en deux parties indépendantes, quel que soit le modèle sous-jacent. Or du fait même du processus de génération des données, leur $\hat{\lambda}$ était systématiquement fortement corrélé aux X_2.

1.3.2 Le modèle en deux équations indépendantes

C'est le modèle préconisé dans le cas où le *sample selection model* est difficilement applicable, notamment dans le cas où le $\hat{\lambda}$ est fortement corrélé aux variables de seconde étape.

Le modèle classique en deux équations

Ce modèle en deux parties vise à estimer les deux équations (1.34) et (1.35) de façon indépendante. Le biais de variable omise (le λ correspondant au biais de sélection) est donc négligé, afin de gagner en erreur quadratique moyenne. Dans le cas des dépenses de santé (ou même en économie du travail), la variable expliquée est souvent spécifiée en *log* (cf section 1.2). Le modèle comprend donc une première équation modélisant la probabilité de recours (autrement dit, la probabilité d'avoir $y > 0$), sous la forme *logit* ou *probit*. La deuxième équation spécifie le *log* de la variable expliquée y comme une fonction linéaire des variables x : $E[ln(y)|y > 0, x] = x\beta$. Les deux équations peuvent être estimés séparément, et les problèmes liés à la retransformation *log* cités section 1.2.1 ressurgissent lors de l'estimation de seconde étape par *MCO*. On néglige le biais de sélection : ce peut être justifié par Leung et Yu (1996) car l'objectif est de minimiser l'erreur quadratique moyenne de l'estimation. Par exemple, Manning, Newhouse, Duan, Keeler, et Leibowitz (1987), utilisant l'expérience de la RAND, proposent un modèle en 4 équations indépendantes :

1. explication de la participation au système de soins,
2. pour les participants, explication de la participation à l'hôpital
3. pour les consommateurs de soins d'hôpital, explication du niveau de dépense totale
4. pour les consommateurs de soins ambulatoires uniquement, explication du niveau de dépense totale

Alors même qu'il pourrait y avoir corrélation entre les perturbations des différentes équations, les auteurs utilisent un modèle en 4 équations indépendantes afin de minimiser l'erreur quadratique moyenne des estimateurs. Dans les équations expliquant le niveau de dépenses de soins, la dépense est exprimée en *log*. Les prédictions de dépense doivent donc être retransformées (voir section 1.2.1) : les auteurs choisissent pour l'hôpital d'utiliser le *homoscedastic smearing estimate* de Duan (voir section 1.2.2) calculé non pas

sur tout l'échantillon, mais par sous-groupe d'individus, dont le point commun est le niveau de remboursement de leur plan d'assurance santé. En effet, le *homoscedastic smearing estimate* est un estimateur du facteur de retransformation Ψ convergent dès lors qu'il ne dépend pas des variables explicatives X du modèle.

Afin d'éviter le problème de la retransformation, Mullahy (1998) a formalisé un modèle en deux parties modifié (*modified two-part model*), qui est devenu une référence dans la littérature en économie de la santé, et particulièrement dans le domaine du vieillissement, modèle que nous utilisons dans le chapitre 2.

Le modèle en deux parties indépendantes modifié

Le *modified two-part model* de Mullahy (1998) consiste d'une part à supposer l'indépendance entre les deux équations du modèle, évitant de ce fait les problèmes liés au *sample selection model* (cités dans la section 1.3.1), d'autre part à employer un *GLM* pour l'équation de seconde étape, évitant ainsi les problèmes liés à la retransformation (voir section 1.2.1). Un avantage de cette méthode est la souplesse apportée par la spécification *GLM* pour l'équation de seconde étape : on peut ainsi traiter indifféremment de variables continues, ou de comptage. Le modèle estime d'abord l'équation de participation sur tout l'échantillon, puis l'équation de consommation conditionnelle sur les consommants uniquement. On a, pour la deuxième étape, $E[y|y > 0, X] = \exp(X\beta)$ avec y suivant une loi de la famille exponentielle de moyenne μ, avec $g(\mu) = X\beta$, g étant la fonction de lien (habituellement le *log*) (Mullahy 1998). Le *GLM* de la seconde équation est un modèle fondé sur une loi dont la moyenne serait μ. Dans le cas de dépenses de santé par exemple, on a donc $\mu > 0$ dans la deuxième équation du modèle, estimée uniquement sur les consommants. Les lois usuellement employées sont les lois de Poisson ou Gamma, dont les supports sont respectivement N et R^+. Le fait que l'on impose $\mu > 0$ revient à prendre en compte dans le *GLM* de la seconde équation une loi à valeurs strictement positives. Dans le cas d'une densité continue comme la loi Gamma, cela ne pose pas de problème : en effet, $Prob(x > 0) = 1$. Mais pour des densités de support différent de R^{+*} ou N^*, il faut les tronquer à zéro. Plus exactement, on considère la loi de y, conditionnelle à l'événement

$[y > 0]$.

Soit une une loi de densité f : on a $f(x|x > 0) = \frac{f(x)}{Prob(x>0)}$. Dans le cas de modèles de comptage, par exemple dans le cas d'une loi de Poisson (de support N) de paramètre μ, on a $Prob(x > 0) = 1 - Prob(x = 0) = 1 - e^{-\mu}$. On a vu que le pseudo-maximum de vraisemblance pouvait employer une loi à support discret pour modéliser une variable continue (cf section 1.2.3). Curieusement toutefois, Seshamani et Gray (2004a et b) n'intègrent pas la troncature à zéro dans les spécifications de la deuxième équation d'un *modified two-part model* employant une loi de Poisson, et ce sans justification.

1.4 Les modèles zero-inflated

Ce type de modèle s'applique à des variables de comptage. L'hypothèse de départ diffère de celle du *(modified) two-part model*, pour qui le grand nombre de zéros dans une distribution provient du phénomène suivant : les individus décident d'abord de consommer ou non, et ceux qui ont décidé de consommer décident ensuite de leur niveau de consommation strictement positif. Le raisonnement sous-jacent aux modèles *zero-inflated* est le suivant : la population se divise en deux types d'individus, ceux qui ne consommeront de toute façon aucun soin, quelle que soit leur situation, et ceux qui sont susceptibles de recourir au système de santé ; la consommation de ces derniers pouvant être nulle. Une illustration souvent évoquée est celle de la pêche : le fait de n'avoir pêché aucun poisson peut provenir du fait d'être allé à la pêche et de n'y avoir rien attrapé, *ou bien* du fait de ne pas être du tout allé à la pêche. Dans le cas de la consommation de soins de santé, la valeur 0 peut donc provenir soit du fait que l'individu appartient au groupe des non-consommants, soit du fait que l'individu appartient aux individus susceptibles de consommer, mais il ne l'a pas fait. Cette modélisation prend explicitement en compte la possible exclusion du système de soins : les gens exclus du système ne consommeront de toute façon pas, même s'ils sont malades. Soit f_1 la densité du processus binaire codant le fait d'être susceptible ou non de consommer (ex : *logit*) et f_2 la densité de probabilité de la consommation de soins (ex : Poisson). La densité f de la variable de consommation de soins y s'écrit :

$$f(y) = f_1(0) + [1 - f_1(0)]f_2(0) \ si \ y = 0$$

$$f(y) = [1 - f_1(0)]f_2(y) \ si \ y > 0$$

L'estimation se fait par maximum de vraisemblance. Ce type de modèle est plus rarement employé que les modèles en deux équations car il prédit trop peu de zéros (Greene 2002, Cameron et Trivedi 2006). Nous ne l'emploierons pas dans les analyses ultérieures.

Chapitre 2

Vieillissement et dépenses de santé : réévaluer l'importance de la démographie

Ce chapitre a fait l'objet d'une publication dans la revue Health Economics (2006, vol. 15) sous le titre "Health expenditure growth : reassessing the threat of aging" [1]

2.1 Introduction

Le vieillissement de la population est souvent considéré comme le facteur principal de l'évolution future des dépenses de santé. En effet, au niveau individuel, le niveau de dépense de santé est une fonction croissante de l'âge (voir graphique 2.1). L'espérance de vie ne cessant de croître dans les pays développés, la proportion et le nombre grandissants de personnes âgées devraient mener quasi-mécaniquement à une croissance accélérée des dépenses de santé.

[1] Dormont, Grignon, et Huber (2006)

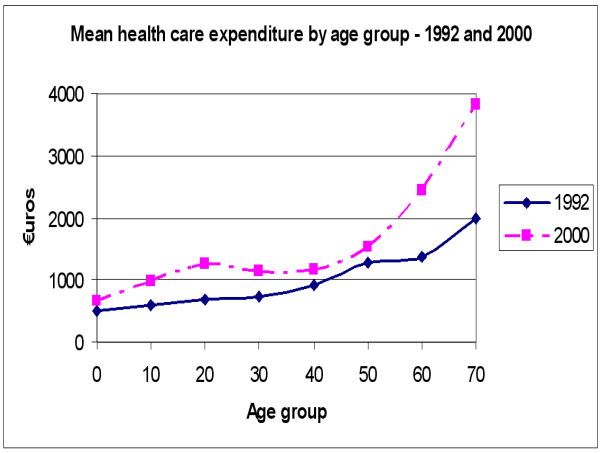


FIG. 2.1: Dépense de santé moyenne par tranche d'âge, années 1992 et 2000, données IRDES

On constate sur le graphique 2.1 que le profil de dépense moyenne par âge n'est pas resté constant entre les années 1992 et 2000, ce qui peut être dû à des changements de morbidité, ou de comportement. En tout état de cause, cet effet est susceptible d'être un facteur explicatif important de la hausse des dépenses de santé observée entre ces deux années, outre les effets des changements démographiques.

Le but de ce chapitre est d'évaluer l'importance du vieillissement démographique dans l'évolution des dépenses de santé observée entre les années 1992 et 2000. La hausse des dépenses de santé en France peut s'expliquer par trois facteurs différents : (*i*) l'effet purement démographique (l'augmentation du nombre et de la proportion des personnes âgées) ; (*ii*) les changements de morbidité à âge donné ; (*iii*) les changements de pratiques, à âge et santé donnés. Ce dernier effet peut être lié au progrès technique et aux changements de comportement. Le but de cette étude est d'identifier, évaluer et interpréter les effets respectifs de ces trois facteurs. Le premier relève de l'effet purement démographique, alors que les deux derniers sont liés aux changements de profil de dépense observés, changement que nous analysons par le biais de micro-simulations.

Nos données sont constituées de deux échantillons représentatifs de 3441 et 5003 individus observés en 1992 et 2000. Il s'agit de l'Appariement, élaboré par l'IRDES. Nous

utilisons des techniques de micro-simulations pour évaluer rétrospectivement les composantes du déplacement de profil observé entre les années 1992 et 2000 sur le graphique 2.1. Nos résultats montrent que les changements de morbidité induisent un déplacement vers le bas du profil, alors que les changements de pratiques induisent un déplacement vers le haut de ce profil de dépense.

Après avoir évalué les influences respectives des changements de morbidité et de pratiques au niveau micro-économique, nous appliquons nos résultats à la structure par âge de la population française, afin de comparer les effets agrégés des changements démographiques et de profils entre les années 1992 et 2000. Au niveau macro-économique, l'augmentation des dépenses de santé due aux changements démographiques est très faible, en comparaison des changements de pratiques. Les résultats montrent que dans la hausse des dépenses totales, l'impact des changements de pratiques est 3,8 fois plus élevé que celui des changements dans la structure par âge de la population.

Nos données présentent l'avantage de fournir des informations non seulement sur le "risque vital", qui est un indicateur de la proximité de la mort, mais également sur de nombreux autres indicateurs de morbidité. De plus, plutôt que de n'étudier que les dépenses d'hôpital (Seshamani et Gray 2004b, Seshamani et Gray 2004a) ou bien les dépenses totales (Zweifel, Felder, et Meiers 1999), nous avons choisi d'étudier séparément les trois composantes principales des dépenses de santé individuelles : les dépenses de médecin, les dépenses de pharmacie soumises à remboursement, ainsi que les dépenses d'hôpital. Cette analyse détaillée nous permet de tenir compte de la nature très différente de ces dépenses. Par exemple, les dépenses de pharmacie et d'hôpital sont très influencées par le progrès technique, au contraire des dépenses de médecin.

Le chapitre est organisé comme suit : dans la première section 2.2 nous décrivons l'apport de la littérature à ce sujet. Puis, en section 2.3, nous présentons les principales caractéristiques des données et discutons de l'accès aux soins des personnes âgées dans le système de santé français. Puis, nous étudions la relation entre morbidité et âge (section 2.4). L'approche empirique est présentée dans la section 2.5. Les résultats micro-économétriques sont présentés dans la section 2.6. Dans la section 2.7, nous comparons les

effets du vieillissement à ceux des changements de pratiques au niveau macro-économique. Nous concluons en section 2.8.

2.2 L'apport de la littérature

De nombreuses études portent sur les conséquences du vieillissement sur la croissance des dépenses de santé. Nous présentons dans cette section les traits saillants de ces études, qui ont toutes pour enjeu d'évaluer l'impact relatif de l'âge sur les dépenses. Nous montrons que tant sur le plan micro que macro-économique, l'identification des différents effets menant à l'augmentation des dépenses se révèle difficile.

2.2.1 L'impact du vieillissement apparaît marginal au niveau macro-économique

Au niveau macro-économique, la grande majorité des travaux de recherche a pour conclusion que l'effet de la structure par âge sur les dépenses de santé est faible, voire insignifiant, alors que l'impact du PIB (ou niveau de vie) est hautement significatif[2]. Une étude de Grignon (2002) a évalué les effets sur la dépense totale des changements démographiques à dépense par âge constante : un tel scénario "naïf" n'évalue l'augmentation de la dépense remboursée (hors soins de long terme) qu'à 0,9 point de PIB entre 1997 et 2020, augmentation "significative mais non explosive"[3]. Récemment, l'OCDE (2006) a publié des décompositions rétrospectives de la hausse des dépenses publiques de santé, ainsi que des projections à 2050. Pour l'ensemble des pays de l'OCDE, l'effet de l'âge représente moins d'un dixième de la hausse des dépenses de santé observée entre 1970 et 2002. Quant aux projections, les auteurs soulignent que des facteurs non démographiques, dont le progrès technique, sont les facteurs principaux de la hausse des dépenses de santé. Pour les pays de l'OCDE, la part des dépenses de santé dans le PIB est prévue d'augmenter de 3,9 points à l'horizon 2050 (de 5,7% en 2005 à 9,6% en 2050), dont seulement 0,7 point

[2](Grignon 2002, Bac et Cornilleau 2002, Gerdtham, Sogaard, Andersson, et Jonsson 1992, Gerdtham, Jonsson, MacFarlan, et Oxley 1998, Getzen 1992, Hitiris et Posnett 1992, Leu 1986, O'Connell 1996, OCDE 1987, Gerdtham et Jönsson 2000)

[3]Pour une dépense totale liée aux remboursements par l'Assurance Maladie représentant $5,5\%$ du PIB en 2000.

serait dû au vieillissement démographique. Dans une étude plus ancienne, L'Horty, Quinet, et Rupprecht (1997) aboutissent à la conclusion que "la progression des dépenses de santé en France sur les 35 dernières années s'explique de façon prépondérante par celle du niveau de vie". L'impact du vieillissement serait largement minoritaire, et n'expliquerait de toute façon pas les différences tendancielles entre pays. Les auteurs soulignent également l'importance du facteur d'offre, qui comprend notamment le système de paiement des praticiens (à l'acte, ou à la capitation) et le progrès technique, qui sont selon eux des facteurs déterminants de l'évolution des dépenses de santé. Les principaux déterminants de la hausse des dépenses de santé sont selon les auteurs le progrès technique (mesuré par le volume des dépenses thérapeutiques par lits d'hôpitaux) et l'évolution du PIB par tête. Le vieillissement (mesuré très simplement par la part des plus de 65 ans dans la population totale) a selon cette étude un impact marginal. Les auteurs rappellent que la relation entre revenu et dépense est beaucoup plus probante au niveau macro-économique qu'au niveau individuel : en effet, les individus sont la plupart du temps assurés, donc bénéficient d'une relative gratuité des soins. Au niveau global se retrouverait la contrainte budgétaire liée au PIB. Pour Gerdtham et Jönsson (2000), le niveau de PIB est le facteur principal expliquant la différence de dépense totale de santé entre les pays. Les auteurs soulignent que l'élasticité-revenu des dépenses de santé apparaît le plus souvent supérieure à 1 : la santé aurait donc les caractéristiques d'un bien de luxe. L'augmentation de la part des dépenses de santé dans le PIB des pays développés en serait une manifestation tangible. Selon Lubitz (2005), cette proportion est passée aux Etats-Unis de $5,7\%$ en 1965 à $14,9\%$ en 2002.

Les études réalisées sur données macro-économiques permettent difficilement d'identifier les différents effets menant à l'augmentation des dépenses de santé. Si le niveau du PIB détermine de façon univoque le niveau de dépenses de santé et si le progrès technique est également entièrement déterminé par le PIB, la relation entre progrès technique et dépenses de santé n'est pas identifiable. Quant à l'effet du vieillissement démographique, il apparaît marginal dans les études macro-économiques, alors que le niveau de dépense de santé est une fonction croissante de l'âge et que les personnes âgées sont de plus en plus

nombreuses.

A l'échelle d'un seul pays, l'identification des mécanismes à l'œuvre dans la détermination des dépenses de santé est donc délicate si l'on se limite au cadre d'analyse macro-économique. Les études sur données individuelles permettent quant à elles de distinguer ces effets.

2.2.2 La question de la proximité de la mort

La question de l'effet de l'âge sur les dépenses de santé a été abordée par de nombreuses études sur données micro-économiques. Ces études se sont largement focalisée sur la question de la proximité de la mort (le *time to death*). Lubitz et Riley (1993) ont posé les jalons de cette approche, en remarquant que les dépenses étaient concentrées lors des dernières années, voire derniers mois de la vie des individus. Pour Zweifel, Felder, et Meiers (1999), la principale explication du niveau de la dépense de santé individuelle est la proximité de la mort, et non l'âge en tant que tel. L'article, intitulé *Ageing of population and health care expenditures : a red herring ?* est frappant, et tout d'abord par son titre. En effet, un "hareng rouge" désigne en Anglais un détail mineur qui attire l'attention, détournant ainsi de choses plus importantes. Pour Zweifel, Felder et Meiers, c'est le vieillissement de la population qui joue ce rôle parmi les facteurs habituellement retenus dans l'explication de l'évolution des dépenses de santé. L'apport de leur étude est de comparer au niveau individuel l'effet de l'âge calendaire à l'effet de la proximité de la mort dans l'explication des dépenses de santé. De données en coupe simple on tire habituellement une forte corrélation entre l'âge des patients et leur dépense de santé. Une réaction naturelle laisserait donc penser que la hausse continue de la proportion des personnes âgées dans la population totale est la source principale de l'augmentation des dépenses de santé. Zweifel et ses co-auteurs démentent cette affirmation en montrant sur données d'assurés suisses de 65 ans et plus que l'âge des individus n'intervient en rien dans l'explication de leurs dépenses de santé.

Pour les auteurs, le lien mis en évidence entre âge et dépenses de santé n'est qu'indirect : plus précisément, les dépenses de santé sont fonction de la proximité du

décès, plutôt que de l'âge en lui-même. Si les personnes de 90 ans coûtent plus cher en moyenne que celles de 80 ans, ce serait donc tout simplement parce que la proportion d'individus vivant leur dernière année de vie est plus importante chez les premiers que chez les seconds. Ces résultats ont été confirmés par de nombreuses études (Seshamani et Gray 2004b, Seshamani et Gray 2004a, O'Neill, Groom, Avery, Boot, et Thornhill 2000, McGrail, Green, Barer, Evans, Hertzmann, et Normand 2000).

En particulier, Seshamani et Gray (2004a et b) s'inspirent de l'étude de Zweifel, Felder, et Meiers (1999) en analysant l'impact de la proximité du décès sur les dépenses de santé sur des données anglaises de malades hospitalisés de plus de 65 ans, dont le décès est observé. Sur des données extrêmement complètes, fournissant le relevé exhaustif des dépenses hospitalières, la nature des interventions (par spécialité) et les déclarations de décès, les auteurs aboutissent à la même conclusion que Zweifel, Felder, et Meiers (1999). L'importance de la proximité de la mort est ainsi confirmée, même en contrôlant par l'état de santé des individus (ce qui n'était pas fait dans Zweifel, Felder, et Meiers (1999)).

Le "coût de la mort" est en effet considérable. L'étude de Lubitz et Riley (1993) a en effet montré qu'entre un quart et un tiers des dépenses totales de *Medicare* étaient engagées par des personnes en dernière année de vie, cette proportion étant stable sur la période d'étude considérée (1976-1988). Les deux derniers mois de vie concentrent quant à eux la moitié des dépenses encourues lors de la dernière année de vie. Les études antérieures par Riley, Lubitz, Prihoba, et Rabey (1987) et Lubitz et Prihoba (1984) aboutissaient à des conclusions similaires.

Yang, Norton, et Stearns (2003) dans une étude descriptive montrent que le profil de dépense totale par âge des bénéficiaires de *Medicare* (cette dépense totale étant principalement déterminée par les dépenses d'hôpital) est effectivement très influencé par les personnes en fin de vie. Examinant le profil de dépense par âge des "survivants", ils trouvent un profil de dépense par âge beaucoup moins pentu : la proximité de la mort est donc pour ces auteurs un facteur essentiel du niveau de dépense par âge. Stearns et Norton (2004) démontrent en effet que dans des projections de dépense de santé prenant

en compte l'augmentation de la longévité mais omettant de prendre en compte le facteur proximité de la mort, les prévisions de dépense sont biaisées vers le haut (biais de 15% pour des prédictions à 2020). Cependant, Yang, Norton, et Stearns (2003) montrent que le surcoût dû à la dernière année de vie décroît avec l'âge du décès (résultat également présent chez Felder, Meier, et Schmitt (2000)) : l'allongement de la durée de vie moyenne est donc susceptible de faire diminuer les dépenses liées à la fin de vie.

2.2.3 La proximité de la mort : simple *proxy* de la morbidité ?

Si dans l'explication des dépenses de santé individuelles le facteur *proximité de la mort* joue effectivement un rôle, soulignons que les études l'ayant mis en évidence sont menées sur des individus de 65 ans et plus. Stooker *et al.*[4] montrent que lorsque l'on considère les dépenses de santé de toute une vie, le coût de la dernière année de vie ne représente que 10% du coût total de soins, relativisant ainsi l'importance de la dernière année de vie sur laquelle beaucoup d'études (largement basées sur des échantillons de personnes déjà âgées) se focalisent. Hogan *et al.*[5] vérifient ainsi sur un échantillon de bénéficiaires de *Medicare* que les personnes en dernière année de vie consomment beaucoup plus que les autres. Cependant, munis d'informations sur les morbidités des individus, ils montrent qu'à morbidité identique, une personne en dernière année de vie ne consomme pas beaucoup plus qu'une autre. Pour les auteurs, les dépenses sont donc principalement dues à la morbidité et non à la proximité de la mort : la corrélation apparente entre proximité de la mort et dépenses de santé ne proviendrait que de la corrélation existant entre morbidité et dépense. Salas et Raftery (2001) avaient critiqué la démarche économétrique de Zweifel, Felder, et Meiers (1999), en mettant en doute l'exogénéité du régresseur *proximité de la mort* dans l'équation expliquant la dépense de santé. Si le montant de soins engagé a une influence sur la proximité de la mort, l'effet de la proximité de la mort sur les dépenses est mal évalué. Zweifel, Felder, et Meiers (2001) ont répondu de manière relativement convaincante, mais il demeure que la proximité de la mort est dans leur étude la seule variable reflétant l'état de santé individuel. En effet, les données dont ils disposent ne fournissent pas d'information précise sur les pathologies individuelles. Dans les études

[4]Stooker, van Acht, van Barneveld, van Vliet, van Hout, Hessing, et Bussbach (2001)
[5]Hogan, Lunney, Gabel, et Lynn (2001)

de Seshamani et Gray (2004a et b), les causes d'hospitalisation sont présentes dans les données et dans l'analyse. Cependant, les auteurs accordent peu d'importance à l'impact de la morbidité individuelle par laquelle ils contrôlent, et se focalisent exclusivement sur l'analyse de l'impact sur les dépenses de la proximité de la mort. Or, la proximité de la mort peut n'être vue que comme un *proxy* de l'état de santé de l'individu. Un *proxy* est en effet une variable ne reflétant qu'imparfaitement le phénomène que l'on cherche à décrire. Dans le cas des dépenses de santé, ce n'est pas la proximité de la mort en tant que telle qui déclenche la dépense, mais plutôt l'état de santé de l'individu. En l'absence d'information sur la gravité de l'état de santé d'un individu, la proximité de la mort n'en donne qu'une idée *ex post*.

La morbidité apparaît ainsi comme le facteur principal expliquant la dépense au niveau individuel. Trois théories coexistent concernant l'évolution de l'état de santé des populations vieillissantes : l'expansion de la morbidité (Kramer 1980), la compression de la morbidité (Fries 1980), et l'équilibre dynamique (Manton 1982). Les données françaises suggèrent une compression de la morbidité. En effet, entre 1980 et 1990, l'espérance de vie à la naissance a augmenté de 2,5 ans pour les deux sexes alors que dans le même temps, l'espérance de vie sans incapacité augmentait de 3 ans pour les hommes et de 2,7 ans pour les femmes (Sermet (1998) citant Robine et Mormiche (1994)). Dans le cas américain, Cutler et Sheiner (2001) montrent que l'augmentation de l'espérance de vie conjuguée à la diminution avec l'âge du coût de la dernière année de vie peut contrecarrer l'augmentation des dépenses de santé liées au vieillissement. Ils soulignent également l'importance de la diminution généralisée de la prévalence d'invalidité, qui mène également à une réduction des dépenses.

Les résultats sur l'espérance de vie sans incapacité suggèrent donc une compression de la morbidité. Cependant, on a assisté entre 1970 et 1991 à une augmentation de la morbidité déclarée (Sermet 1998). Ces deux phénomènes ne sont pas forcément contradictoires : l'augmentation de la morbidité déclarée peut provenir d'une meilleure information des malades (outre l'augmentation du nombre de malades, dû au vieillissement), tandis que le progrès technique peut avoir permis une meilleure prise en charge de ces maladies.

Regardant les dépenses de soins de ville, Mizrahi et Sabourdin (1997) montrent qu'il existe un fort effet générationnel : à âge donné, chaque génération consomme plus que la génération précédente. Alors que jusqu'aux années 1980, la dépense de santé moyenne diminuait à partir de 80 ans, la dépense augmente continûment avec l'âge dès les années 90, phénomène que l'on retrouve sur nos données (dont la première vague d'enquête disponible date de 1992).

2.2.4 Le progrès technique

Cutler et Sheiner (2001) montrent que le progrès technique, en augmentant les coûts, contrecarre largement les effets d'amélioration de la morbidité, et constitue de fait un fardeau difficilement soutenable. Lubitz (2005) souligne que l'allongement de la durée de vie permis par le progrès technique n'est pas susceptible de mener à une réduction des dépenses de santé. L'auteur, évoquant le modèle *FEM* (*Future Elderly Model*) de la Rand[6], montre que les coûts liés au progrès technique annulent la réduction des dépenses liée à l'amélioration de l'état de santé des personnes âgées. L'allongement de la durée de vie est en effet synonyme d'augmentation des dépenses, liée d'une part au plus grand nombre d'années durant lesquelles ces personnes consomment, et d'autre part à l'apparition massive de coûts liés aux soins de long terme. L'auteur souligne néanmoins que ces simulations sont faites sur des personnes éligibles à *Medicare* : une amélioration de la santé des individus à des âges plus jeunes est susceptible de mener à une réduction des dépenses de ces personnes une fois devenus âgées. Cutler (2005) souligne toutefois que les prédictions du *FEM* sont établies en supposant que les innovations augmentent systématiquement les coûts de traitement. Le progrès technique peut cependant réduire les coûts d'un traitement : la non prise en compte de cet effet peut ainsi mener à surestimer les coûts dûs au progrès technique.

L'objet de ce chapitre est de présenter une méthode novatrice permettant d'identifier simplement les effets d'âge, de progrès technique, et de changements d'état de santé sur l'évolution des dépenses, au niveau micro puis macro-économique.

[6]Ce modèle prédit l'évolution des niveaux de dépense pour les bénéficiaires de *Medicare* grâce à des cohortes d'individus dont on simule le parcours de santé à travers des micro-simulations. Le progrès technique est introduit dans les simulations au travers d'innovations augmentant les coûts de traitement.

2.3 Les données

2.3.1 Le système de santé français

Le système de santé français est public et universel. Environ 99% de la population est couverte, sans restriction d'âge : la couverture est continue sur toute la durée de la vie. Durant la période considérée (1992 à 2000), le système d'assurance a octroyé un accès illimité aux soins, sans restrictions de fréquence. Ce système couvre presque 100% des dépenses de soins d'hôpital et 70% des dépenses de soins ambulatoires. Les individus ont la possibilité de souscrire à une assurance complémentaire. Cette décision peut émaner de l'individu, mais c'est souvent l'employeur qui fournit cette couverture. La population française souscrit à 80% à une complémentaire, qui couvre les 30% restant à charge pour les dépenses ambulatoires. En France, la valeur totale des prescriptions de médicaments augmente rapidement du fait des prix toujours plus élevés des innovations mises sur le marché. Au total, alors que le nombre de consultations a augmenté de 18% entre 1992 et 2000, les dépenses de pharmacie ont augmenté de 59% (IRDES 2002).

Grâce à ce niveau de couverture, les personnes âgées françaises ne sont pas confrontées à des contraintes financières qui pourraient limiter leur accès aux soins. De plus, comme la couverture est la même tout au long de la vie, il n'y a pas de raison de constater un "saut" dans la consommation de soins à un âge particulier. Ce saut pourrait être observé dans le cas américain car les personnes sont incitées à reporter les soins jusqu'à la date où elles peuvent être prises en charge par *Medicare*, c'est-à-dire 65 ans. Le système d'assurance santé français n'a donc aucune raison d'influencer le profil de dépenses de santé par âge.

La période étudiée est caractérisée par la modification des pratiques médicales, notamment concernant les personnes âgées. Par exemple, les recommandations visant à prévenir l'hypercholestérolémie ont été récemment étendues en France aux âges très avancés. De la même manière, l'utilisation de la chirurgie de la cataracte a drastiquement changé pour les personnes âgées en France : pour les 75-84 ans, la proportion de traitement chirurgical de la cataracte est passée de 40% à 55% entre 1993 et 1998 (Baubeau, Bousquet, et Joubert 2001). Un autre exemple est l'utilisation de plus en plus répandue de

l'angioplastie pour le traitement de l'infarctus du myocarde. En France, la proportion de traitement par angioplastie est passée pour les 75-84 ans de 9,6% en 1993 à 22% en 2001 (Oberlin, Mouquet, et Folliguet 2004). Du côté de la demande, des changements sur les préférences accordant plus d'importance au bien-être des personnes âgées peuvent expliquer l'utilisation accrue de procédures innovantes. Du côté de l'offre, le progrès technique améliore les résultats tout en améliorant la sécurité des interventions, ce qui permet l'extension de procédures innovantes aux personnes les plus âgées. Ceci peut mener à des coûts plus élevés, mais également à un bien-être plus élevé des personnes âgées.

2.3.2 Les données

Nous utilisons les données de l'Appariement, construit par l'IRDES. Elles concernent 3441 assurés français en 1992 et 5003 pour l'année 2000. Ces données sont des données d'enquête (SPS, i.e. *Santé Protection Sociale*), enquête conduite par l'IRDES[7] sur un sous-échantillon de l'EPAS (i.e. *Echantillon Permanent d'Assurés Sociaux*). L'EPAS est un échantillon d'individus de la Caisse d'Assurance Maladie du Régime Général pour lesquels toutes les dépenses soumises à remboursement sont recueillies. Nous nous limitons ici aux membres de la CNAMTS (Caisse Nationale d'Assurance Maladie des Travailleurs Salariés) car seule cette caisse faisait l'objet d'une telle enquête en 1992[8]. On a donc les caractéristiques individuelles et les recueils de consommation d'un échantillon d'adhérents de la CNAMTS (travailleurs salariés, salariés retraités, et leurs familles).

L'échantillon fournit une information fiable sur les dépenses individuelles de santé, l'adhésion éventuelle à une couverture complémentaire, la morbidité, et les caractéristiques socio-démographiques comme l'âge, la catégorie socio-professionnelle et le revenu net. L'échantillon est composé d'individus vivant en ménage ordinaire. L'étude ne traite donc pas des soins de long terme, qui concernent principalement les personnes vivant en institution.

[7] Anciennement CREDES, Institut de Recherche et de Documentation en Economie de la Santé, 10, rue Vauvenargues, 75018 Paris, France

[8] La CNAMTS couvre environ 83% de la population (Sandier, Polton, Paris, et Thomson 2002)

Nous nous focaliserons sur trois sous-groupes de dépenses : les dépenses de consultations, les dépenses pharmaceutiques liées aux soins ambulatoires, et les dépenses d'hôpital. Les dépenses de médecins et de pharmacie peuvent être facilement recueillies individuellement par la Caisse d'Assurance Maladie, à travers les demandes de remboursement qui lui sont adressées. Ce n'est pas le cas des soins hospitaliers. En effet, les hôpitaux français sont dans leur grande majorité publics. Durant la période étudiée, ils sont financés sur la base d'un budget global. De ce fait, les dépenses de santé individuelles ne sont pas observées directement : une évaluation est calculée par l'Assurance-Maladie sur la base de la durée du séjour et du coût moyen journalier de l'unité de soins en question, avec un degré de détail assez fin concernant la nomenclature des unités de soin.

Afin de rendre les dépenses de santé comparables entre les années 1992 et 2000, nous avons corrigé les dépenses de l'inflation et converti toutes les valeurs en Euros. L'indice des prix[9] est spécifique au type de dépense considéré. Entre 1992 et 2000, l'indice des prix des consultations de médecins a augmenté de 9,15%, reflétant les changements de tarifs, qui sont régulés. L'indice des prix des dépenses pharmaceutiques ambulatoires a augmenté de 2,04%. Cet indice ne reflète pas les prix des innovations apparues sur le marché, car chaque innovation est conçue comme un changement de produit. L'indice des prix des dépenses d'hôpital a augmenté de 16,54%. Comme pour l'indice des prix des dépenses pharmaceutiques, cet indice ne reflète pas le prix des innovations. En effet, le coût moyen de séjour en unité de soin est calculé sur la base de la quantité de travail fournie par les médecins et infirmières (durée et intensité), et l'indice des prix est calculé suivant une combinaison de l'indice des salaires des fonctionnaires et des biens de consommation courante. Ainsi, nous ajustons simplement par le taux général d'inflation : la croissance des dépenses de santé est toujours influencée par le progrès technique, auquel nous nous intéressons ici.

2.3.3 Analyse descriptive

Le tableau 2.1 présente les caractéristiques générales de l'échantillon. Les nombres en italiques mentionnés dans le tableau précisent les valeurs correspondantes trouvées dans

[9] Source : DREES et Ecosanté (IRDES 2002)

les comptes de la santé. La proportion de chaque type de dépense dans notre échantillon est ainsi relativement proche de ce que l'on observe sur les comptes de la santé. Nous avons choisi de ne pas éliminer les *outliers* de l'analyse, car ôter de l'échantillon les gros consommateurs de soins rendrait les données peu représentatives de la réalité (en particulier sur les dépenses d'hôpital).

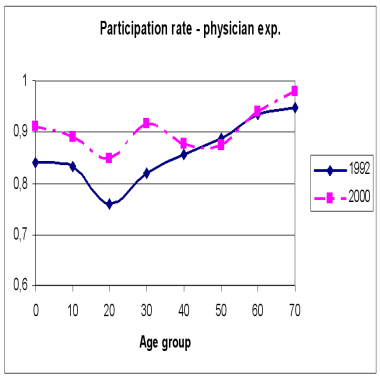
	1992	2000
N	3441	5003
% femmes	54,37%	50,19%
Dépense de santé moyenne par tête (euro courant)	874,1	1386,81
Taux de participation	86,60%	92,14%
Structure des dépenses de santé		
Proportion des dépenses de consultations	33,86%	25,81% *(22,5)*
Proportion des dépenses de pharmacie	23,23%	27,02% *(22,1)*
Proportion des dépenses d'hôpital	42,91%	47,17% *(41,5)*
Croissance (période 1992-2000)		
Dépenses de consultations		7,75% *(28,1)*
Dépenses de pharmacie		64,41% *(51,7)*
Dépenses d'hôpital		55,35% *(28,0)*

Tab. 2.1: Caractéristiques de l'échantillon

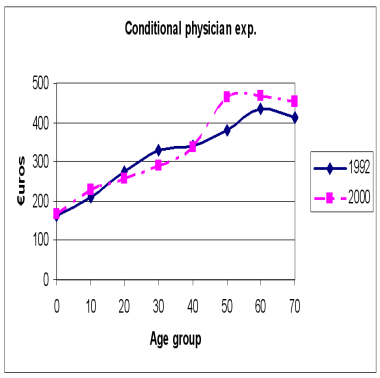
Une caractéristique des dépenses de santé que l'on retrouve sur nos données est une proportion non négligeable de non consommants. Schématiquement, le niveau de dépense est expliqué par le taux de participation ainsi que par la dépense conditionnelle à la participation. Les graphiques (2.2a) à (2.2i) présentent, pour chaque type de dépense considérée (consultations, pharmacie, hôpital) le profil par âge du taux de participation, de la dépense conditionnelle, et de la dépense non conditionnelle. Les taux de participation pour les consultations et la pharmacie sont élevés (autour de 90%). Les changements observés dans les profils de participation ont donc un impact réduit sur la dépense non conditionnelle. Pour ces dépenses, les consommations conditionnelles et inconditionnelles sont très proches (voir par exemple les graphiques (2.2e) et (2.2f)), ce que l'on ne retrouve pas pour l'hôpital. En effet, pour les dépenses d'hôpital, la participation est faible : entre 10% et 20%, selon la tranche d'âge considérée. Les changements de participation ont ainsi un impact plus fort sur les changements de profil de la dépense non conditionnelle. Entre

1992 et 2000, le déplacement du profil de la dépense non conditionnelle est conséquent pour les personnes de 50 ans et plus. On voit ici clairement que les dépenses ambulatoires et hospitalières ont des caractéristiques dissemblables, qui nécessiteront des analyses séparées.

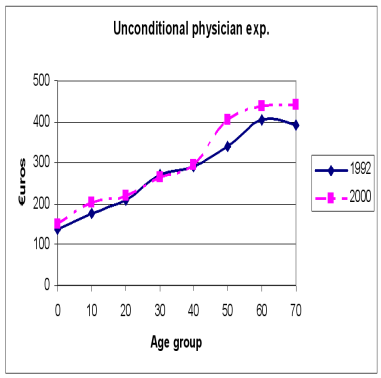
FIG. 2.2: Profils de participation et de dépense



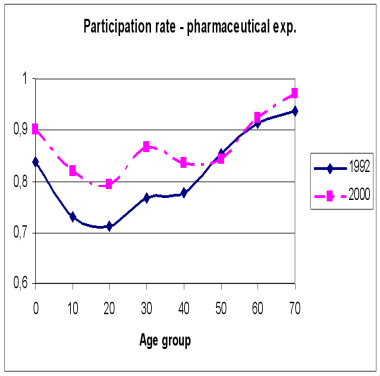
(2.2a) 0, 10, 20, 30
p=0,0000



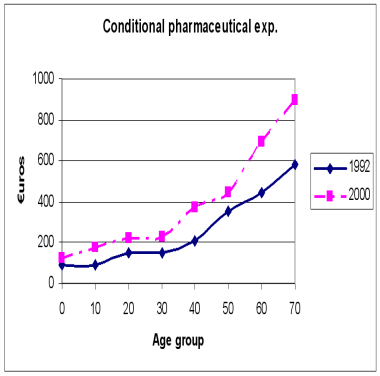
(2.2b) 50
p=0,1201



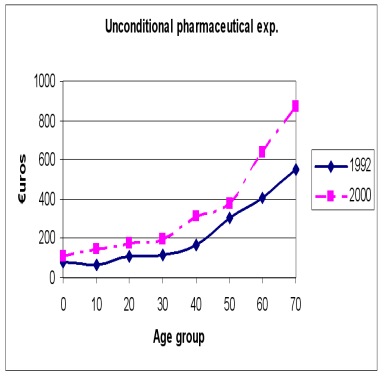
(2.2c) 50
p=0,2253



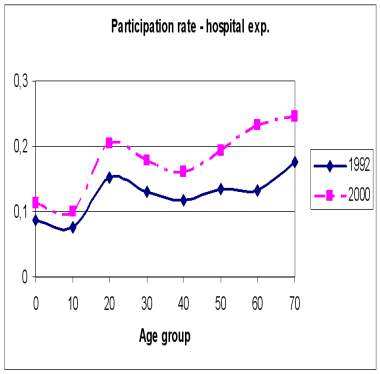
(2.2d) 0, 10, 20, 30, 40
p=0,0000



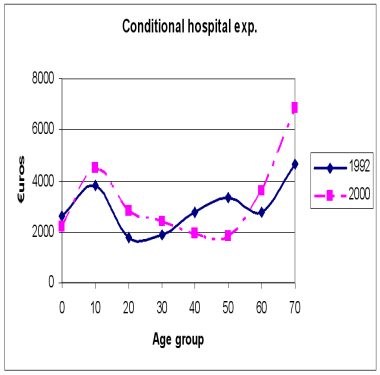
(2.2e) 40, 60, 70
p=0,0000



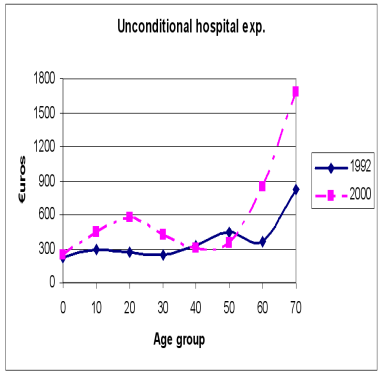
(2.2f) 10, 30, 40, 60, 70
p=0,0000



(2.2g) 20, 30, 40, 50, 60
p=0,0000



(2.2h) néant
p=0,5370



(2.2i) 60, 70
p=0,0057

Nous avons mis en œuvre des procédures de test de la significativité de la différence entre les profils de dépense des deux années. Les résultats sont fournis sous chaque graphique : nous donnons les groupes d'âge pour lesquels la différence de dépense moyenne

est significative entre 1992 et 2000, ainsi que la p-value du test de significativité global de différence entre les deux profils.

Pour les dépenses de consultations et de pharmacie, la dépense moyenne (conditionnelle et inconditionnelle) augmente avec l'âge : le profil de dépense est croissant avec l'âge. Un déplacement de ce profil vers le haut apparaît particulièrement marqué pour les personnes de 40 ans et plus, entre 1992 et 2000. Ce déplacement est significatif pour les dépenses de pharmacie (voir (2.2f) : $p < 0,0001$) alors qu'il apparaît non significatif pour les dépenses de consultations (voir (2.2c) : $p = 0,22$).

Pour la dépense d'hôpital (conditionnelle et inconditionnelle), un déplacement vers le haut du profil est visible, mais uniquement pour les personnes de 60 ans et plus (graphique (2.2i)). Le taux de participation pour les soins d'hôpital a augmenté de manière significative entre 1992 et 2000, pour tous les groupes d'âge. Cette augmentation de la participation explique le déplacement de profil de la dépense inconditionnelle entre les deux années : ce déplacement est significatif (p<0,0001, ce qui est dû aux groupes d'âge 60-69 ans et 70 ans et plus), alors que le déplacement du profil des dépenses conditionnelles n'est pas significatif (p=0,54).

Le graphique concernant les dépenses inconditionnelles d'hôpital peut appuyer les arguments de Zweifel, Felder, et Meiers (1999) et Seshamani et Gray (2004a et b), à savoir que la proximité de la mort est le principal facteur expliquant le niveau des dépenses d'hôpital. En effet, le profil de dépenses est relativement plat jusqu'au groupe d'âge 50-59 ans, puis devient très pentu (graphique (2.2i)). Yang, Norton, et Stearns (2003) montrent que le profil par âge des dépenses de *Medicare* (dont le total est largement influencé par les dépenses d'hôpital) est fortement influencé par les dépenses engendrées par les personnes en fin de vie, qui décèdent dans l'année d'observation. En éliminant de l'échantillon les personnes en fin de vie, les auteurs montrent que le profil par âge des dépenses de santé est nettement moins pentu. Il est donc essentiel de vérifier que le profil pentu que nous observons sur nos données quant aux dépenses d'hôpital n'est pas uniquement dû aux personnes

en fin de vie. L'enquête IRDES répliquée sur le même échantillon à 4 ans d'intervalle souffre d'attrition comme toute enquête, mais permet d'identifier un grand nombre de personnes ayant répondu à l'enquête en 2000 et en 2004. Le graphique 2.3 présent ci-dessous montre les profils de dépense de pharmacie et d'hôpital pour tous les individus de l'année 2000, ainsi que pour les individus de l'année 2000 dont nous savons qu'ils ont survécu jusqu'en 2004. La pente du profil est moins prononcée, mais toujours visible pour ces "survivants". On montre ainsi que la forme du profil des dépenses d'hôpital n'est pas uniquement dû aux personnes en fin de vie. Remarquons que le profil par âge des dépenses de pharmacie est le même pour les "survivants" et pour tout l'échantillon. Il est intéressant de remarquer que le profil de dépenses par âge ne croît qu'à partir de la tranche d'âge 60-69 ans, alors que dans le cas des dépenses pharmaceutiques, ce profil est croissant dès la tranche d'âge 0-9 ans, et ce même pour les survivants.

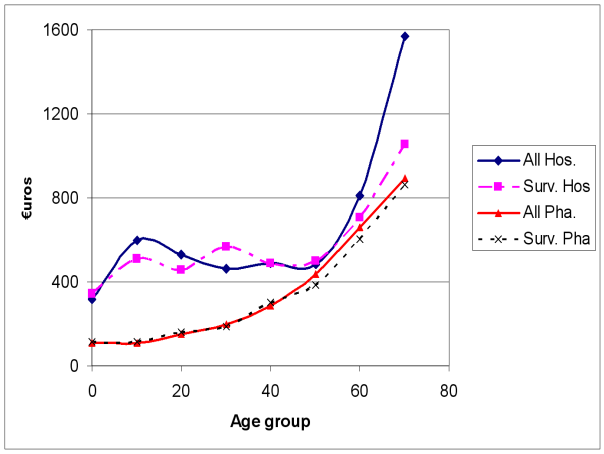


FIG. 2.3: Dépenses de pharmacie et d'hôpital (année 2000), tous et "survivants"

2.4 Lien entre morbidité et âge

La construction d'indices de morbidité pertinents a constitué une part importante de l'analyse. En effet, le niveau de morbidité par groupe d'âge et ses changements à travers le temps jouent un rôle central dans nos évaluations empiriques. Les données fournissent une information détaillée à propos de l'état de santé de chaque individu, ainsi que la liste des maladies dont cet individu peut souffrir. De plus, de nombreuses autres caractéristiques individuelles sont recueillies : la taille, l'indice de masse corporelle (IMC), le fait d'être fumeur, ainsi que des variable socio-économiques comme le statut d'emploi (employé, chômeur, inactif, retraité), et le fait d'être ou non bénéficiaire du RMI. Aucune de ces variables n'est utilisée comme régresseur dans les équations de participation et de consommation conditionnelle. Elles peuvent donc être utilisées comme des instruments pour tester l'exogénéité des indicateurs de morbidité.

L'information dont nous disposons est la plupart du temps déclarative, à part les relevés objectifs de consommation de soins. Lors de la construction des bases de données par l'IRDES, des médecins vérifient l'enquête et corrigent les incohérences entre les informations recueillies (par exemple, comparaison entre les médicaments consommés et les pathologies déclarées lors de l'enquête). Des indicateurs synthétiques de morbidité sont construits par l'IRDES, comme le nombre de maladies, le degré d'invalidité et le degré de risque vital. Cette dernière variable est une note donnée par le médecin qui code l'enquête, relative à la probabilité de décès dans les 5 ans. Cette variable possède 6 niveaux, allant de 0 (risque zéro) à 5 (pronostic sûrement mauvais). Outre ces indicateurs présents dans l'enquête, nous avons construit des indicateurs pour la présence de maladies chroniques comme l'hypertension, le diabète, etc (en tout, 10 variables). Tous les indicateurs sont décrits en annexe B.1.3. Le processus de sélection des variables et leur procédure de test sont également décrits en annexe B.2, menant à la liste finale d'indicateurs retenus. La construction et la sélection des indicateurs de morbidité ont été le fruit d'un travail relativement long, mais permettent de dépasser l'approche standard en termes de *proximité de la mort* (*time to death*) par une mesure assez fine de la morbidité. Notons que beaucoup

d'études ne prennent en compte que la *proximité de la mort*, et celles qui prétendent tenir compte de la morbidité ne disposent en général que du degré d'invalidité.

2.4.1 La sélection d'indicateurs de morbidité exogènes

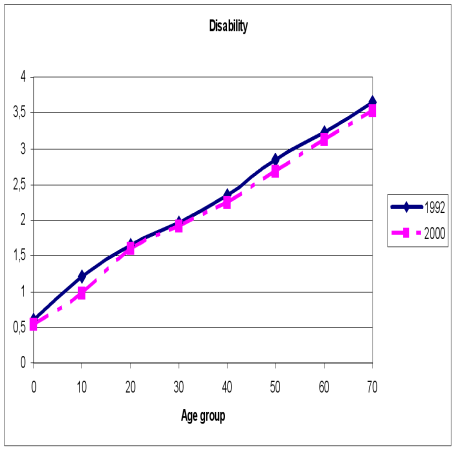
Si les soins de santé remplissent leur office, l'état de santé n'est pas un régresseur exogène dans une équation expliquant les dépenses de santé. Son exogénéité dépend du délai entre la dépense encourue et les améliorations éventuelles de l'état de santé. Le fait que nos données soient composées de deux coupes couvrant chacune les dépenses d'une année a des conséquences sur l'exogénéité des régresseurs : les maladies chroniques sont susceptibles d'être exogènes. En effet, les soins médicaux ne peuvent guérir une maladie chronique mais peuvent retarder ou limiter ses effets délétères. Les maladies les plus répandues (diabète, hypertension artérielle, cardiopathies) ne peuvent être guéries, et leur survenance est indépendante du montant de soins accordé à l'individu. En revanche, les procédures comme le remplacement d'une hanche par une prothèse améliore substantiellement les conditions de vie du patient : le degré d'invalidité peut être non-exogène.

La plupart des analyses consacrées à l'impact du vieillissement sur les dépenses de santé ont intégré l'état de santé comme une variable de contrôle dans les estimations, sans aucun moyen de vérifier si ces variables utilisées pour mesurer l'état de santé étaient bien exogènes. Grâce à la richesse de nos données, qui rendent disponibles des instruments pertinents, nous avons pu procéder à des tests d'exogénéité pour les variables suivantes : degré d'invalidité, risque vital, note sur l'état de santé auto-attribuée, et nombre de maladies. Les indicateurs synthétiques de morbidité sont en effet susceptibles d'être non-exogènes, car durant l'année d'observation, les soins engagés peuvent avoir un impact sur ces indicateurs. La procédure utilisée est décrite dans la section 2.6. En revanche, nous avons supposé que les maladies chroniques étaient exogènes, pour les raisons cités plus haut.

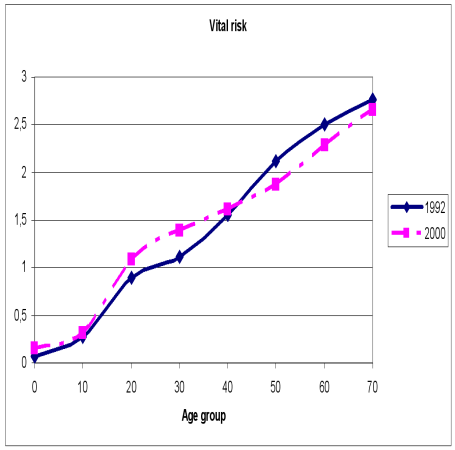
2.4.2 Le lien étroit entre la morbidité et le profil de dépenses par âge

Il est bien connu que la morbidité augmente avec l'âge. Cela apparaît clairement sur les graphiques suivants (2.4), qui montrent, pour 4 indicateurs de morbidité, leur niveau moyen (ou prévalence moyenne) par tranche d'âge.

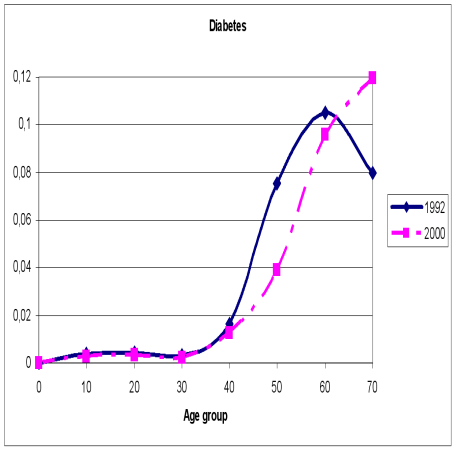
FIG. 2.4: Evolution d'indicateurs moyens de morbidité par tranche d'âge



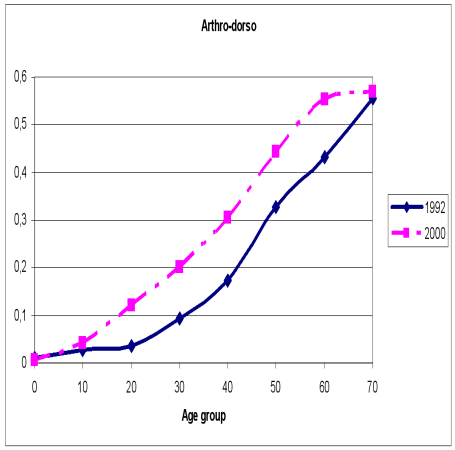
(2.4a) Invalidité



(2.4b) Risque vital



(2.4c) Diabète



(2.4d) Arthrite, arthropathies, dorsopathies

Une analyse par moindres carrés ordinaires, estimés sur données non transformées, montre que le fait que les dépenses de santé augmentent avec l'âge est entièrement dû à l'augmentation de la morbidité avec l'âge. Le graphique suivant présente la contribution au profil de dépense pharmaceutique par âge de chacun des 3 groupes de variables suivants : des indicatrices d'âge, des indicateurs de morbidité, ainsi que des indicateurs socio-économiques (sexe, revenu, couverture complémentaire, catégorie socio-professionnelle, niveau d'études)[10]. Le graphique (2.5) montre clairement que les indicatrices d'âge n'influencent pas la forme du profil de dépenses par âge, une fois la morbidité

[10]Ceci a été obtenu pour l'année 2000. Des graphiques comparables sont obtenus pour l'année 1992, et pour les autres types de dépenses. Des résultats similaires sont également obtenus pour le taux de participation : le profil croissant avec l'âge du recours aux soins est expliqué principalement par l'évolution de la morbidité avec l'âge

prise en compte : ce sont bien les indicateurs de morbidité qui déterminent la forme du profil de dépense par âge.

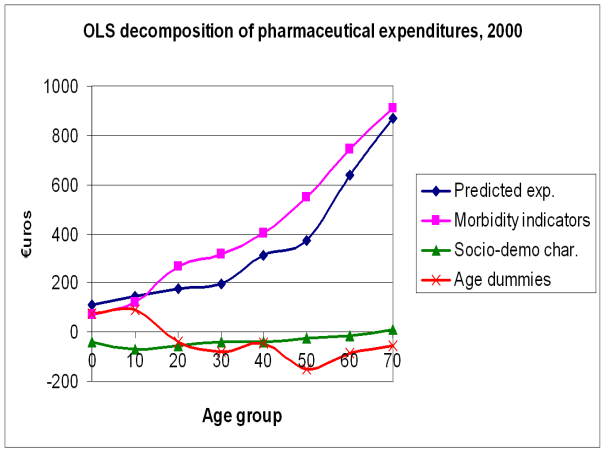


FIG. 2.5: Effet des différents prédicteurs de la dépense pharmaceutique, 2000

2.4.3 Les changements de morbidité à travers le temps

Les graphiques (2.4) illustrent aussi les changements de morbidité par groupe d'âge qui sont survenus entre les années 1992 et 2000. Pour alléger la présentation, nous n'avons pas détaillé les courbes pour chacun des autres indicateurs. Ces graphiques sont représentatifs de ce qui est trouvé pour nos divers indicateurs. Pour la plupart d'entre eux, on constate une diminution de la morbidité entre les deux années, autrement dit une amélioration de l'état de santé. Cependant, ce mouvement n'est pas systématique, et non homogène pour tous les âges. Par exemple, la prévalence du diabète diminue pour tous les groupes d'âge entre 1992 et 2000, sauf pour les individus âgés de 70 ans et plus (graphique (2.4c)). A l'inverse, la prévalence des arthrites, arthropathies et dorsopathies augmente pour chaque groupe d'âge (graphique (2.4d)).

Ces résultats ne permettent pas de tirer de conclusion générale sur une éventuelle amélioration de l'état de santé entre 1992 et 2000 : le niveau de prévalence des indicateurs de morbidité n'a pas d'évolution uniforme selon l'indicateur considéré, ni selon le groupe d'âge considéré. L'aspect novateur de notre approche est de permettre d'évaluer l'effet total de tous ces changements sur les dépenses moyennes par tranche d'âge. Autrement dit, il

est possible de synthétiser l'impact total de ces changements de morbidité sur le niveau de dépense de santé. Anticipant sur nos résultats, nous pouvons dès à présent révéler que cet impact total est négatif, signifiant que l'état de santé des individus s'est globalement amélioré entre 1992 et 2000.

2.5 L'approche empirique

Le but de nos simulations est d'évaluer l'influence de différents effets sur le déplacement du profil par âge des dépenses de santé, observé entre 1992 et 2000. Notre approche empirique comporte trois étapes : (*i*) la spécification et l'estimation d'un modèle en deux équations expliquant la décision de consommer puis le niveau de dépense, conditionnel à la participation (*ii*) l'utilisation des estimations obtenues pour simuler des niveaux moyens contrefactuels de participation et de dépense, par groupe d'âge. Cette deuxième étape permet d'évaluer l'impact des changements de morbidité et de changements de pratiques survenus entre 1992 et 2000. (*iii*) nous utilisons les résultats des micro-simulations pour évaluer au niveau macro-économique les effets respectifs des changements démographiques (le vieillissement) et des déplacements de profil (changements de morbidité et de pratiques) sur la hausse totale des dépenses de santé.

Cette méthode est employée pour chacune des trois composantes principales des dépenses de santé, que nous analysons ici : les dépenses de consultations, les dépenses de pharmacie, et les dépenses d'hôpital.

2.5.1 Spécification économétrique et estimation

Une caractéristique saillante des dépenses de santé est que l'on observe une quantité non négligeable de zéros : beaucoup d'individus n'ont aucune dépense de santé durant la période d'observation. L'analyse descriptive a montré la forte proportion d'individus n'ayant aucune dépense d'hôpital. Pour ce qui est des consultations et des dépenses de pharmacie, la proportion d'individus n'ayant aucune dépense, bien que beaucoup plus faible, est non négligeable. Une telle configuration nécessite des techniques d'estimation particulières. Un grand nombre d'articles sont consacrés au choix entre le *sample-selection model* ou *Heckit* (Heckman 1979) et le *two-part model* (Dow et Norton 2002, Leung et Yu 1996, Manning, Duan, et Rogers 1987)[11]. Leung et Yu (1996) ont montré que les performances du *sample-selection model* dépendent de façon cruciale du degré de colinéarité entre l'inverse du ratio de Mill et les variables explicatives de seconde étape. Lorsqu'il n'y

[11] Les critères de choix entre ces modèles et leurs implications sont exposés en détail au chapitre 1

a pas de colinéarité, un test de Student du coefficient de l'inverse du ratio de Mill peut être utilisé pour choisir entre les deux spécifications. En revanche, si des problèmes de colinéarité surviennent, le *two-part model* est plus fiable car ses performances sont meilleures que le *sample-selection model* en terme d'erreur quadratique moyenne.

Nos données sont caractérisées par une forte corrélation entre l'inverse du ratio de Mill et les variables explicatives de seconde étape : le coefficient de corrélation est situé entre 0,85 et 0,87, selon le type de dépense de santé concerné. Nous avons donc choisi le *two-part model*, qui présente un autre avantage : il permet l'utilisation d'un *GLM* (*generalized linear model*) pour estimer le niveau de dépense conditionnel à la participation. Les *GLM* permettent de prendre en compte assez facilement certaines caractéristiques des dépenses de santé, comme la dissymétrie (*skewness*) de la variable dépense non transformée, les distributions à queue épaisse, et les résidus hétéroscédastiques. En effet, l'approche par *GLM* permet (*i*) d'éviter les difficultés liées à la retransformation des variables exprimées en *log* (voir chapitre 1) (*ii*) de prendre en compte les distributions à queues épaisses et les erreurs hétéroscédastiques, en prenant comme lois de la variable expliquée des distributions de type Poisson ou Gamma (Manning et Mullahy 2001).

Soit l'individu i appartenant au groupe d'âge j. Soit P_{ij} la variable dichotomique codant la participation et C_{ij} la consommation de soins. Considérons le modèle en deux équations :

$$P_{ij} = \mathbf{I}_{P^*_{ij}>0} \quad \text{avec} \quad P^*_{ij} = W'_{ij}c + M'_{ij}b + a_j + u_{ij} = X'_{1,ij}d + u_{ij} \ , \tag{2.1}$$

où $u_{ij} \sim N(0, \sigma^2)$.

$$\begin{cases} C_{ij} = \mathbf{I}_{(P_{ij}=1)} \cdot \left[C^*_{ij}\right] \\ \text{avec } E(C^*_{ij} \Big| X'_{2,ij}) = \exp(Z'_{ij}\gamma + M'_{ij}\beta + \alpha_j) = \exp(X'_{2,ij}\delta) \end{cases} \tag{2.2}$$

Nous avons décidé d'utiliser pour (2.2) un lien *log* : $E(C^*_{ij} \Big| X'_{2,ij})$ est défini comme une fonction exponentielle de $X'_{2,ij}\delta$. En effet, les observations strictement positives des dépenses de santé sont fortement dissymétriques (la *skewness* varie entre 4,50 et 15,47, selon l'année et le type de dépense considérés). Prendre le *log* des dépenses réduit la *skewness* à des valeurs situées entre -0,26 et 0,01. De plus, nous avons utilisé pour C^*_{ij} une distribution Gamma. Ce choix est fondé sur les résultats de tests de Park que nous avons menés, en suivant l'approche suggérée par Manning et Mullahy (2001) (voir chapitre 1).

L'équation (2.1) décrit la décision de faire appel au système de soins et l'équation (2.2) décrit le niveau de consommation. $X'_{1,ij}$ et $X'_{2,ij}$ sont les variables explicatives des équations respectivement de participation et de consommation conditionnelle. Ces régresseurs incluent des indicatrices d'âge a_j et α_j relatives aux groupes d'âge, et des indicateurs de morbidité M'_{ij}. De plus, les équations (2.1) et (2.2) comprennent des variables explicatives relatives aux caractéristiques socio-économiques de l'individu (respectivement W'_{ij} et Z'_{ij}). La liste des indicateurs de morbidité potentiellement utilisés dans les estimations comprend l'invalidité, le risque vital, le nombre de maladies, la note auto-attribuée sur l'état de santé, et des indicatrices codant pour la présence des pathologies suivantes : diabète, broncho-pneumopathies chroniques obstructives et emphysème, cardiopathies ischémiques, hypertension artérielle, maladies des artères, troubles du métabolisme des lipides, dépression, troubles du sommeil, cataracte, et arthrites, arthropathies et dorsopathies. La liste des indicateurs socio-économiques potentiellement utilisés comprend le niveau de revenu, la catégorie socio-professionnelle, le niveau d'études, la présence d'une couverture complémentaire, le sexe, la taille du ménage, et le statut matrimonial. La liste effective dépend de l'équation considérée ((2.1) ou (2.2)) et du type de dépense étudié (dépenses de consultations, de pharmacie, ou d'hôpital) : les variables pertinentes sont sélectionnées grâce à un processus détaillé dans l'annexe B.2.

Les maladies chroniques sont vraisemblablement exogènes. En revanche, les indicateurs synthétiques d'état de santé comme le degré d'invalidité, le degré de risque vital ou le nombre de maladies peuvent être non-exogènes. De plus, la possession d'une couverture

complémentaire peut également être non exogène, car elle peut résulter (du moins en partie) d'une décision individuelle. Nous avons choisi de conserver uniquement les régresseurs exogènes, afin de ne pas ajouter de difficulté méthodologique à une approche relativement complexe.

Grâce à la richesse de nos données, nous avons pu effectuer des tests d'Hausman pour les dépenses de consultations, de pharmacie et d'hôpital. Ceci n'a été possible que sur l'échantillon de l'année 2000. Nous ne disposions pas d'un nombre suffisant d'instruments pour l'échantillon de 1992, où l'enquête n'était pas aussi riche qu'en 2000. Néanmoins, il apparaît légitime de supposer que l'exogénéité vérifiée sur l'échantillon de 2000 le serait également sur l'échantillon de 1992. Soulignons que dans les estimations, l'élimination des variables non exogènes n'introduit pas de biais de variable omise : les tests d'exogénéité sont réalisés de telle sorte que l'exogénéité des variables retenues est vérifiée.

Nous avons utilisé des variables instrumentales pour construire un test d'Hausman, en suivant l'approche de Rivers et Vuong (1988) pour l'équation de participation, et un modèle log-linéaire pour l'équation de consommation conditionnelle. Nous avons utilisé un test de Sargan pour vérifier la validité des instruments utilisés. De plus, nous avons vérifié que cette procédure n'était pas remise en cause par un problème d'instruments faibles (Staiger et Stock 1997). Pour cela, nous avons testé la significativité globale des instruments dans un système d'équation ($SURE$[12]) comprenant autant d'équations que de variables instrumentées, où chaque variable instrumentée est expliquée par les instruments et les régresseurs exogènes. La liste détaillée des instruments et les résultats des tests sont fournis en annexe B.2. Les résultats montrent que la corrélation partielle entre les instruments et les indicateurs de morbidité est hautement significative, avec des statistiques élevées et des degrés de significativité inférieurs à 10^{-3}. Des tests de Sargan ont validé l'exogénéité des instruments. La couverture par une complémentaire apparaît exogène pour les consultations et la dépense pharmaceutique (elle ne constitue pas un régresseur significatif pour l'hôpital). L'exogénéité est rejetée pour le degré d'invalidité dans

[12]Seemingly UnRelated Equations

le cadre des dépenses de consultation. Elle est également rejetée pour le risque vital dans le cadre des dépenses pharmaceutiques.

2.5.2 Prédictions et simulations

Pour l'individu i du groupe d'âge j, la décision de recourir aux services de santé est facilement prédite par l'estimation de (2.1) par un estimateur Probit :

$$\hat{E}(P_{ij} | X_{1,ij}) = \Phi(X'_{1,ij} \hat{d}) \ , \tag{2.3}$$

avec $\Phi(.)$ la fonction de répartition de la loi normale standard.

Pour ce qui est de la dépense conditionnelle, la spécification *GLM* mène à une estimation directe de l'espérance de la dépense conditionnelle, à l'échelle non transformée :

$$\hat{E}(C_{ij} | P_{ij} = 1; X_{2,ij}) = \exp(X'_{2,ij} \hat{\delta}). \tag{2.4}$$

Ce modèle est estimé séparément pour les années 1992 et 2000 pour chaque type de dépense (consultations, pharmacie, hôpital), menant aux coefficients estimés $\hat{d}_{92}$, $\hat{d}_{00}$, $\hat{\delta}_{92}$ et $\hat{\delta}_{00}$.

Comme le calcul des prédicteurs (2.3) et (2.4) fait appel à plusieurs fonctions non-linéaires, nous ne pouvons calculer d'effets additifs. Nous utilisons à la place une approche incrémentale. Pour ce qui est de la décision de participation, nous calculons et simulons sous différentes hypothèses la probabilité prédite de participer, pour le patient moyen de chaque groupe d'âge j. Ceci est fait pour chaque type de dépense (consultations, pharmacie, hôpital).

Pour la population observée en 1992, la probabilité prédite de faire appel au système de soins est la suivante, pour tout individu i appartenant au groupe d'âge j :

$$\hat{\pi}_{ij}^{92.92} = \Phi\left(X'^{92}_{1,ij} \hat{d}_{92}\right) = \Phi\left(W'^{92}_{1,ij} \hat{c}_{92} + M'^{92}_{ij} \hat{b}_{92} + \hat{a}_{j,92}\right)$$

Cette probabilité est évaluée au point moyen de chaque groupe d'âge :

$$\hat{\pi}_{.j}^{92.92} = \Phi\left(X_{1,.j}'^{92}\ \hat{d}_{92}\right) = \Phi\left(W_{1,.j}'^{92}\ \hat{c}_{92} + M_{.j}'^{92}\ \hat{b}_{92} + \hat{a}_{j,92}\right) \tag{2.5}$$

Nous avons choisi cette approche, et non celle qui consisterait à calculer la moyenne par groupe d'âge des probabilités prédites. En effet, les simulations contrefactuelles ne peuvent se faire qu'au niveau du groupe d'âge, et non au niveau individuel, car les individus observés dans les différentes tranches d'âge ne sont pas les mêmes en 1992 et en 2000.

L'effet sur la participation des changements de pratiques à morbidité donnée est mis en évidence en remplaçant $\hat{b}_{92}$ par $\hat{b}_{00}$ dans l'expression (2.5) :

$$\hat{\pi}_{.j}^{m92.b00} = \Phi\left(W_{1,.j}'^{92}\ \hat{c}_{92} + M_{.j}'^{92}\ \hat{b}_{00} + \hat{a}_{j,92}\right) \tag{2.6}$$

L'effet *incrémental* des changements de morbidité survenus entre 1992 et 2000 est mis en évidence en remplaçant, pour chaque groupe d'âge, le niveau moyen de morbidité observé en 1992 $M_{.j}'^{92}$ par le niveau moyen de morbidité observé en 2000, $M_{.j}'^{00}$:

$$\hat{\pi}_{.j}^{m00.b00} = \Phi\left(W_{1,.j}'^{92}\ \hat{c}_{92} + M_{.j}'^{00}\ \hat{b}_{00} + \hat{a}_{j,92}\right) \tag{2.7}$$

Finalement, l'effet incrémental des autres changements de comportement et de caractéristiques individuelles mènent à la probabilité prédite pour l'année 2000 :

$$\hat{\pi}_{.j}^{00.00} = \Phi\left(X_{1,.j}'^{00}\ \hat{d}_{00}\right) = \Phi\left(W_{1,.j}'^{00}\ \hat{c}_{00} + M_{.j}'^{00}\ \hat{b}_{00} + \hat{a}_{j,00}\right) \tag{2.8}$$

Nous avons utilisé l'expression (2.4) pour calculer de façon similaire les prédictions des niveaux de dépense par groupe d'âge, conditionnelle à la participation : $(C|\,P)_{.j}^{92.92}$, $(C|\,P)_{.j}^{m92.b00}$, $(C|\,P)_{.j}^{m00.b00}$ et $(C|\,P)_{.j}^{00.00}$. Plus précisément, les moyennes par groupe d'âge des variables explicatives sont calculées sur le sous-échantillon des participants, en simulant les dépenses conditionnelles, ce qui mène à différents profils de dépense

par âge simulés. Nous avons donc pour la consommation conditionnelle les transitions suivantes :

- La transition de $(C|\,P)_{.j}^{92.92}$ à $(C|\,P)_{.j}^{m92.b00}$ donne l'effet des changements de pratiques à morbidité donnée
- La transition de $(C|\,P)_{.j}^{m92.b00}$ à $(C|\,P)_{.j}^{m00.b00}$ donne l'effet incrémental des changements de morbidité (parmi les participants).
- Finalement, la transition de $(C|\,P)_{.j}^{m00.b00}$ à $(C|\,P)_{.j}^{00.00}$ donne l'effet incrémental des autres changements de comportements et de caractéristiques individuelles survenus entre 1992 et 2000.

A chaque étape de la simulation, la dépense inconditionnelle (donc celle qui serait réellement observée) C est calculée comme le produit $\pi * (C|P)$.

Globalement, les changements de pratiques sont représentés par les changements dans les paramètres estimés de (2.1) ou (2.2) survenus entre 1992 et 2000. Ces changements sont liés aux changements dans le comportement des patients ou des médecins face à une morbidité donnée. Ils résultent également du progrès technique, qui pousse à l'utilisation de médicaments et procédures plus innovants, mais aussi plus coûteux. Comme expliqué dans la section 2.3, ce progrès technique peut concerner l'extension aux âges élevés des procédures innovantes et des protocoles de prévention.

Les changements de pratiques à morbidité donnée sont ainsi capturés par les vecteurs de paramètres estimés $\hat{b}$ et $\hat{\beta}$. Ces paramètres estimés mesurent l'influence de la morbidité sur la participation et la dépense conditionnelle. Ce que nous appelons ici les "autres changements de comportements et de caractéristiques individuelles" survenus entre 1992 et 2000 peuvent également être interprétés comme des changements de pratiques. En fait, ces "autres changements" proviennent des changements dans les variables W et Z, ainsi que des changements dans les paramètres $\hat{c}$, $\hat{\gamma}$, $\hat{a}_j$ et $\hat{\alpha}_j$. Afin de rendre les graphiques relatifs aux micro-simulations lisibles, nous n'avons pas représenté séparément les effets des changements dans les indicatrices d'âge. L'information est cependant présente dans le tableau récapitulatif 2.2 de la section 2.7, où les évaluations sont faites au niveau

macro-économique. Ces changements dans les indicatrices d'âge peuvent être liées à des changements de pratique à morbidité donnée inobservée, mais aussi à des changements dans la morbidité inobservée.

2.6 Résultats micro-économétriques

2.6.1 Estimations

Comme il a été mentionné dans la section précédente, les variables explicatives des équations de participation et de dépense conditionnelles ont été sélectionnées de façon rigoureuse. Seules ont été conservées pour les estimations et simulations celles qui étaient à la fois significatives et exogènes.

L'estimation des équations (2.1) et (2.2) révèle une forte influence de la morbidité sur la participation et la consommation conditionnelle[13]. Les indicateurs de pathologies, le nombre de maladies et les degrés d'invalidité et de risque vital ont un fort impact positif sur le recours aux soins de santé. Par exemple, en 1992, un degré d'invalidité de 3 provoque toutes choses égales par ailleurs une augmentation de la consommation conditionnelle de pharmacie de 42%. Un degré d'invalidité de 4 ou 5 provoque toutes choses égales par ailleurs[14] une augmentation de la consommation conditionnelle de respectivement 77 % et 89 %. La consommation conditionnelle de pharmacie est également fortement influencée par la présence d'hypertension artérielle (+40 %), de diabète (+ 48 %), de troubles du métabolisme des lipides (+ 25 %) et de dépression (+ 37 %)[15]. La participation aux consultations et à la consommation de pharmacie est également significativement influencée par les indicateurs de morbidité. La consommation conditionnelle de consultation est influencée positivement par la dépression (+27 %) et le nombre de maladies. Pour ce qui est des dépenses d'hôpital, la participation apparaît influencée positivement par la présence de diabète, de cataracte, le risque vital, et le niveau de dépense conditionnelle par le degré d'invalidité et l'hypertension artérielle (+37 %).

[13]Tous les résultats des estimations ne sont pas fournis ici (il y a en tout 12 estimations). En annexe B.3 figurent les estimations pour la dépense de pharmacie, pour les deux années

[14]Un niveau d'invalidité de 3 correspond à une situation où l'individu éprouve des difficultés mais vit normalement. Un niveau 4 correspond à une situation où l'individu "doit diminuer son activité domestique ou professionnelle"; un niveau 5 regroupe les situations suivantes : "activité diminuée", "pas d'autonomie domestique" et "confiné au lit".

[15]Ces résultats correspondent aux estimations de 1992

Concernant les autres déterminants de la consommation de soins, les résultats montrent que l'absence de couverture complémentaire réduit significativement la participation et la consommation de soins ambulatoires. Par exemple, en 1992, l'écart (toute choses égales par ailleurs) est de l'ordre de -39% pour la dépense conditionnelle de consultations, et de -25% pour la dépense conditionnelle de pharmacie. En revanche, la couverture complémentaire n'a pas d'influence significative sur le recours à l'hôpital. De manière générale, le sexe n'est pas significatif pour ce qui est de la participation, mais être une femme augmente les dépenses conditionnelles de consultations et de pharmacie d'environ 17%. Par ailleurs, le sexe n'est pas significatif pour les dépenses d'hôpital.

La plupart des coefficients estimés des indicateurs de morbidité sont plus élevés en 2000 qu'en 1992. Cela montre clairement que les changements de pratiques induisent une augmentation des dépenses entre 1992 et 2000, à morbidité donnée.

2.6.2 Les simulations

Les simulations mises en œuvre mènent au calcul de probabilités de participation et de consommation conditionnelles pour l'individu moyen de chaque groupe d'âge j, et ce pour chaque étape du raisonnement. Les valeurs $\hat{\pi}_{.j}^{92.92}$, $\hat{\pi}_{.j}^{m92.b00}$, $\hat{\pi}_{.j}^{m00.b00}$ et $\hat{\pi}_{.j}^{00.00}$ sont calculées pour la participation et $(C|P)_{.j}^{92.92}$, $(C|P)_{.j}^{m92.b00}$, $(C|P)_{.j}^{m00.b00}$ et $(C|P)_{.j}^{00.00}$ pour la dépense conditionnelle. A chaque étape de la simulation, la consommation inconditionnelle C est calculée comme le produit $\pi * (C|P)$. Les profils de dépense par groupe d'âge qui résultent de ces simulations sont exposés dans les graphiques (2.6a) à (2.6i) (les graphiques les plus intéressants sont grossis à la page suivante, sous les appellations (2.7) et (2.8). Pour chaque variable d'intérêt, le déplacement du profil de 1992 (profil 1), au profil de 2000 (profil 4) est décomposé en 3 effets incrémentaux :

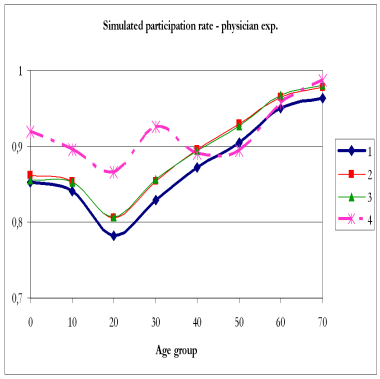
- la transition du profil 1 au profil 2 donne l'effet des changements de pratiques pour un niveau donné de morbidité
- la transition du profil 2 au profil 3 donne l'effet des changements de morbidité
- la transition du profil 3 au profil 4 donne l'effet des autres changements de comportement et de caractéristiques individuelles.

Les soins ambulatoires (consultations et dépenses pharmaceutiques) et d'hôpital mènent à des résultats assez différents, qui méritent un traitement séparé.

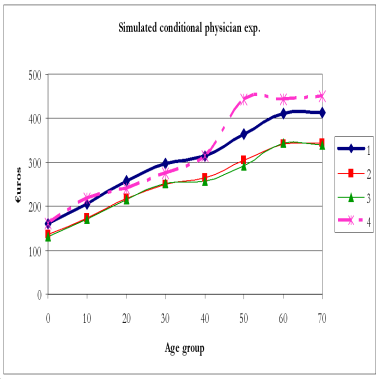
Figure 2.6 : légende des profils

- Profil 1 - dépense prédite, 1992
- Profil 2 - simulation contrefactuelle : dépense prédite avec les caractéristiques de la population de 1992 et les coefficients de morbidité de 2000
- Profil 3 - simulation contrefactuelle : dépense prédite avec les caractéristiques de la population de 1992, les coefficients de morbidité de 2000, et le niveau de morbidité de 2000
- Profil 4 - dépense prédite, 2000.

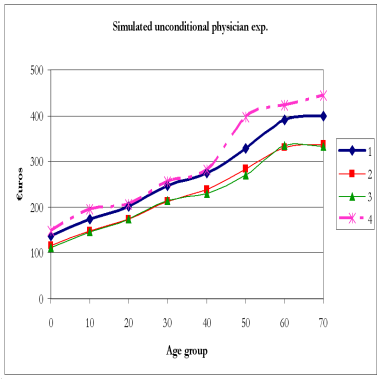
FIG. 2.6: Profils simulés



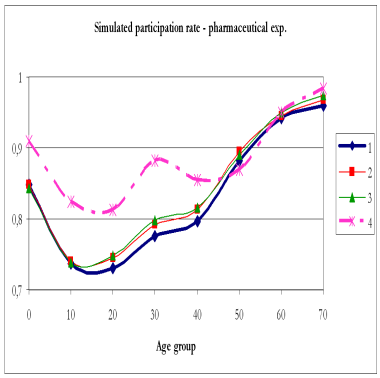
(2.6a)



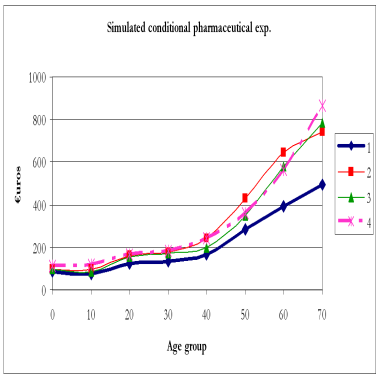
(2.6b)



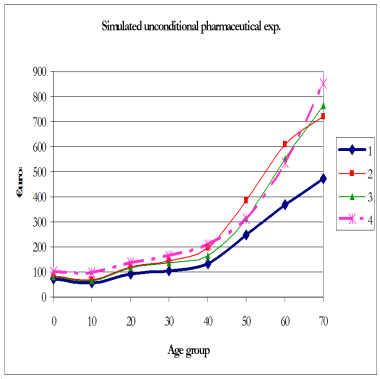
(2.6c)



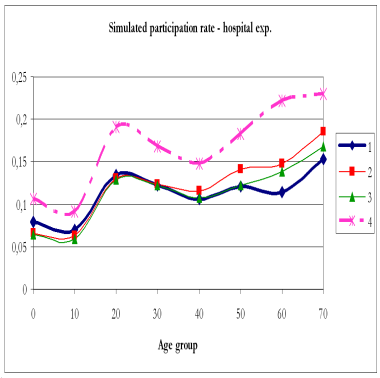
(2.6d)



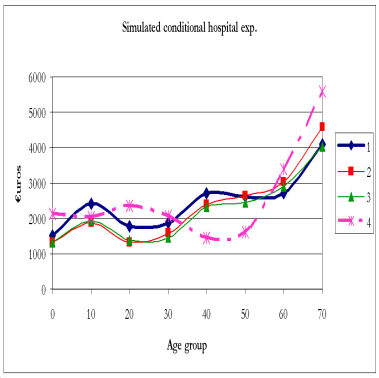
(2.6e)



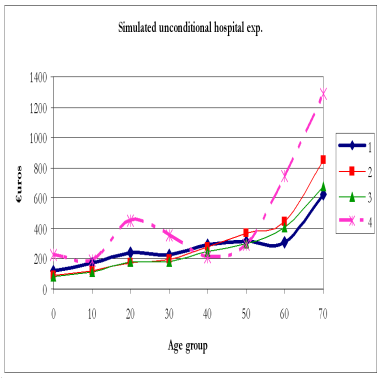
(2.6f)



(2.6g)



(2.6h)



(2.6i)

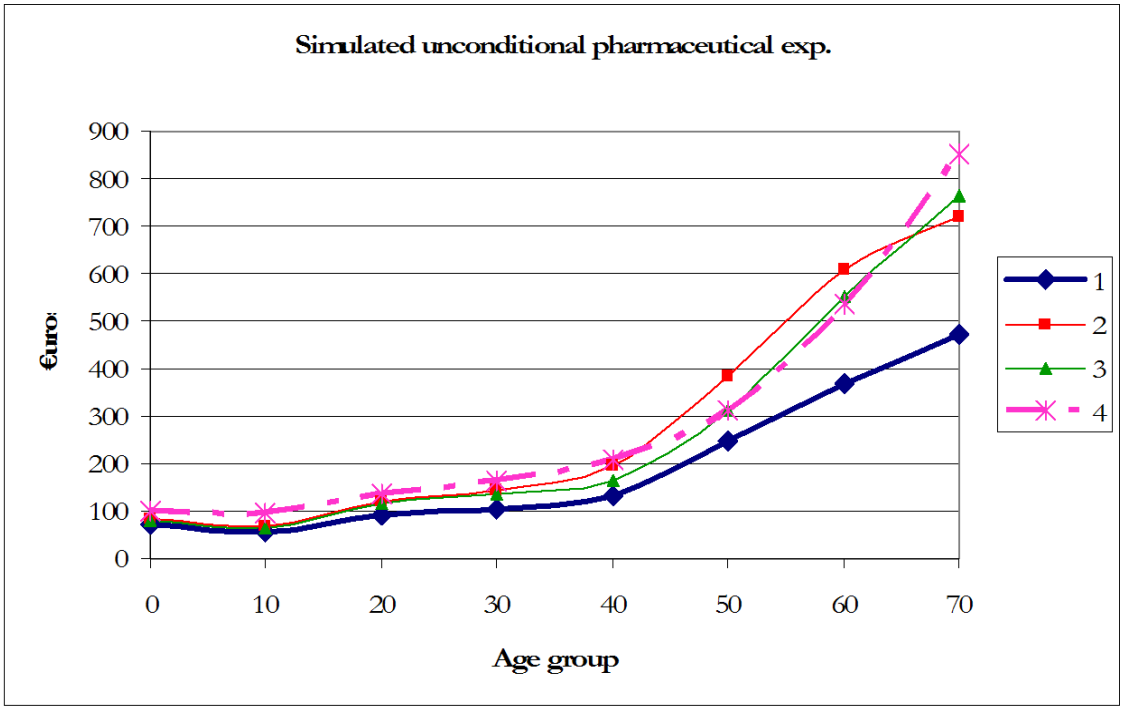


FIG. 2.7: Graphique 2.6f (agrandi) : Dépense pharmaceutique inconditionnelle simulée

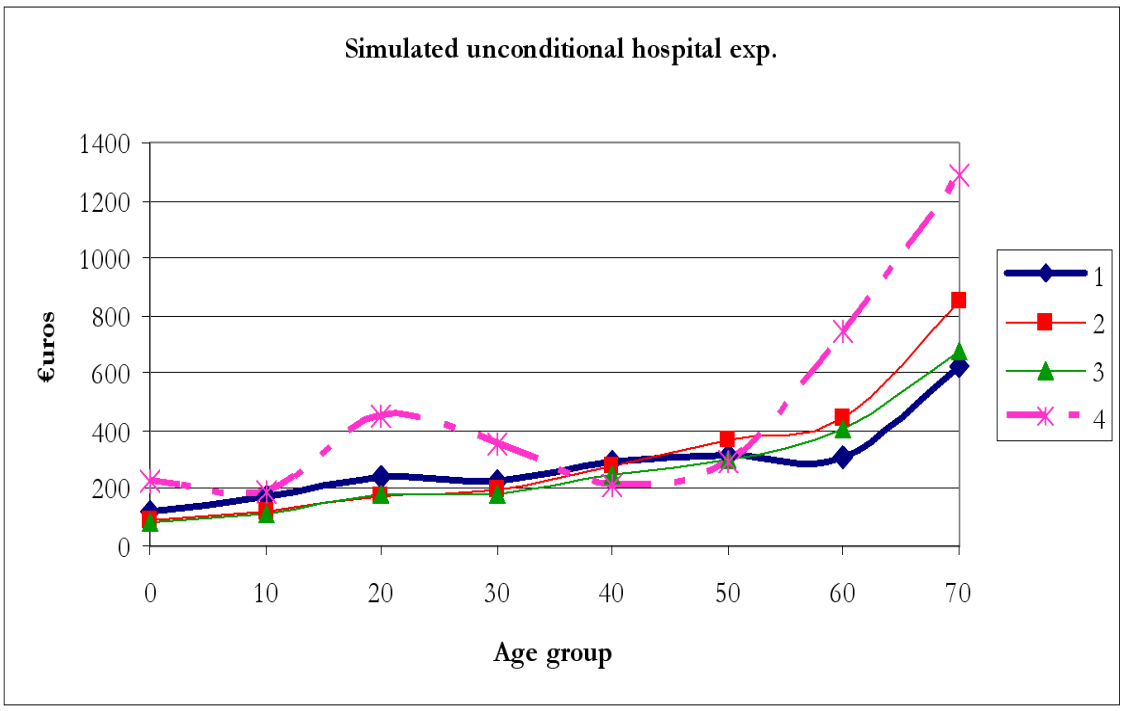


FIG. 2.8: Graphique 2.6i (agrandi) : Dépense hospitalière inconditionnelle simulée

Les soins ambulatoires

Pour les soins ambulatoires, la majorité des changements observés sont dûs aux changements dans la consommation conditionnelle, et non aux changements de participation. Les changements de taux de participation ont une influence très faible sur le déplacement des profils de consommation inconditionnelle. En effet, la participation aux soins ambulatoires est déjà très élevée en 1992 (entre 80% et 98% selon les tranches d'âge considérées), et elle varie peu entre les deux années : l'impact de ces changements sur la consommation observée est donc nécessairement limité.

Le déplacement du profil de dépenses par âge est beaucoup plus spectaculaire pour les dépenses de pharmacie qu'il ne l'est pour les dépenses de consultation. On a vu d'ailleurs que le déplacement de profil pour les dépenses de consultation n'est pas significatif[16]. Cette nette distinction entre les dépenses pharmaceutiques et celles de consultations apparaît en comparant les graphiques (2.6c) et (2.6f) (tout en tenant compte des différences d'échelle).

Concernant les dépenses de consultations, on observe un faible déplacement vers le bas du profil, dû aux changements de pratiques (graphique (2.6c), profil 1 à 2) : à morbidité donnée, les individus consomment moins de dépenses de consultations en 2000 qu'en 1992. Un déplacement vers le haut, qui est principalement dû aux "autres changements", compense quasi totalement cet impact négatif : au total, les changements survenus entre 1992 et 2000 ne sont pas significatifs pour les dépenses de consultations.

Les dépenses de médicaments sont à l'origine de la plus grande partie de l'augmentation des dépenses de santé totales. Les changements pour chaque groupe d'âge sont clairement observables sur le graphique (2.7) (version agrandie de (2.6f)). Nos simulations montrent que cette augmentation est presqu'entièrement due à un important déplacement du profil vers le haut dû aux changements de pratiques à morbidité observable donnée (profil 1 à 2). Les changements de morbidité induisent un déplacement du profil vers le bas, qui

[16] Voir les tests sur les graphiques (2.2b) et (2.2c)

ne compense que partiellement le déplacement vers le haut observé pour les changements de pratiques. Pour les dépenses non conditionnelles de pharmacie, le déplacement de profil est dû au déplacement de profil de la dépense conditionnelle et non à celui du profil de la participation. De plus, le déplacement de profil est non significatif pour les dépenses de consultations. Ces résultats suggèrent la présence d'une forte composante innovation dans ces changements : le déplacement vers le haut du profil n'est pas dû aux changements de comportement des personnes âgées, mais plutôt à l'offre de nouveaux produits sur le marché de la santé.

L'hôpital

Ce qui est observé pour l'hôpital contraste sensiblement avec les résultats trouvés pour les soins ambulatoires, au sens où les changements dans le profil de dépense totale sont principalement dûs aux changements dans le taux de participation par groupe d'âge (voir graphique (2.8), version agrandie du (2.6g)). La hausse du taux de participation est observée pour tous les groupes d'âge, et son amplitude augmente avec l'âge. Nos simulations montrent que ce déplacement vers le haut du profil est dû par ordre d'importance croissante aux "autres changements" et aux changements de pratiques à morbidité donnée. Le déplacement du profil vers le haut dû aux changements de pratiques (profil 1 à 2) est visible pour les 50 ans et plus, ainsi qu'un léger déplacement vers le bas dû aux changements de morbidité pour ces mêmes groupes d'âge.

Concernant les dépenses non conditionnelles d'hôpital, nous trouvons de même que précédemment un effet négatif de changements de morbidité, pour les personnes de 40 ans et plus. Nous trouvons également un effet positif des changements de pratiques à morbidité donnée. Pour les deux dernières tranches d'âge (60-69 ans et 70 ans et plus), on observe un déplacement vers le haut très prononcé du profil, dû aux effets combinés des changements de pratiques et des "autres changements". Ce déplacement n'est visible que sur le graphique (2.6i) (dépense non conditionnelle), car il est dû aux changements dans les comportements de participation.

2.7 Evaluer les effets comparés du vieillissement et des changements de pratiques au niveau macro-économique

Dans le tableau 2.2, nous appliquons les prédictions et simulations calculées en section 2.6.2 à la population française totale observée en 1992 et 2000. Ceci nous permet d'évaluer rétrospectivement les effets respectifs des changements démographiques et des changements de profils de dépense survenus entre 1992 et 2000. De plus, la décomposition des changements de profils permet d'affiner l'analyse en distinguant l'apport des différentes composantes des changements de profil au niveau macro-économique.

L'effet du changement démographique total entre 1992 et 2000 est calculé en supposant que la dépense moyenne par âge est gardée constante à son niveau de 1992, et que seule la taille et la structure de la population varient[17]. Les autres effets sont ensuite calculés à population de 2000 constante, en faisant varier les différents profils de dépense simulés. Une analyse bootstrap a été menée afin d'évaluer la significativité des résultats (bootstrap par paires, 800 réplications).

Nous trouvons encore l'effet négatif des changements de morbidité. L'augmentation des dépenses due aux changements démographiques apparaît faible, comparée à celle due aux changements de pratiques. Pour la dépense totale (comprenant les trois agrégats étudiés : dépenses de consultations, de pharmacie, d'hôpital), nous trouvons que l'impact des changements de pratiques est égal à +12,9%. Cet effet est donc 3,8 fois plus élevé que l'augmentation des dépenses liée aux changements dans la structure d'âge de la population (+3,4%). De plus, nous trouvons que l'impact des changements de morbidité est négatif, suggérant une amélioration de l'état de santé des individus à âge donné. Ceci est vérifié pour la dépense totale, mais également pour chaque type de dépense pris séparément. Pour la dépense totale, cet effet négatif des changements de morbidité (-9,7%) fait plus que compenser le choc positif dû aux changements dans la structure par âge de la population (+3,4%). Ceci est également vérifié pour les dépenses de pharmacie et d'hôpital. Le

[17] Les effectifs de population en question sont présentés dans le tableau B.3 en annexe.

vieillissement n'est alors pas synonyme de dégradation de l'état de santé.

Variation 1992-2000 (%)	Consult.	Pharmacie	Hôpital	Total
Changement démographique total	5,05	7,63	6,37	6,35
dont :				
part due au changement structurel	2,10	4,61	3,38	3,36
part due à l'effet taille	2,95	3,02	2,99	2,99
Changements de pratiques à morbidité donnée	-15,09	52,24	7,06	12,87
Changements de morbidité	-1,23	-9,24	-14,61	-9,74
Changement des indicatrices d'âge	28,63	14,11	-26,25	-1,55
Autres changements	-2,80	2,53	95,51	45,95
Variation totale 1992-2000 (%)	14,57	67,27	68,08	53,89

Tab. 2.2: Simulations au niveau macro-économique

2.8 Conclusion

Nos résultats montrent que le vieillissement n'explique qu'une faible part de l'augmentation des dépenses de santé. L'explication de l'augmentation constatée des dépenses de santé est surtout liée aux changements de pratiques à morbidité donnée. Pour les soins ambulatoires, les changements de pratiques pour les personnes âgées ne sont pas liés à des changements dans les comportements de participation, mais à des changements dans la consommation conditionnelle. Le déplacement de profil dû aux changements de pratiques est spectaculaire pour les dépenses de pharmacie, ce qui suggère une forte composante innovation dans les changements observés. Ces résultats révèlent que les déplacements de profil observés ne sont pas dûs à des changements de comportements du côté de la demande. Ils sont plutôt liés au côté de l'offre, c'est-à-dire à l'introduction de nouveaux produits sur le marché de la santé. La possibilité d'étudier séparément trois grands types de dépenses (consultations, pharmacie, hôpital) a permis de montrer que l'augmentation des dépenses de pharmacie est principalement due à l'offre de nouveaux produits, c'est-à-dire au progrès technique. Prendre en compte la participation dans notre modèle a permis d'isoler des changements qui peuvent être dus à l'initiative du patient. Nos résultats suggèrent que ces changements sont d'importance négligeable.

L'information fournie par l'enquête sur la morbidité individuelle a permis également (*i*) de montrer que l'état de santé s'améliore globalement, y-compris chez les personnes âgées (*ii*) d'évaluer les économies liées à cette amélioration. Les dépenses de santé sont souvent exclusivement conçues comme un coût, et il est en effet très difficile de mesurer leur impact bénéfique en termes d'amélioration de l'état de santé. Nos résultats suggèrent une amélioration de l'état de santé entre 1992 et 2000[18].

Nous avons montré que les changements de pratiques apparaissent comme le facteur principal de la hausse des dépenses de santé constatée entre 1992 et 2000. Ce progrès technique est susceptible d'être orienté particulièrement vers les personnes les

[18] Nos résultats ne prouvent cependant pas que ce sont effectivement les dépenses de santé qui ont mené à une amélioration de l'état de santé, et non d'autres facteurs (hygiène, conjoncture ...).

plus âgées. Dans ce cas, l'impact des changements de pratiques devrait augmenter avec l'âge. Les données, limitées, ne nous ont pas permis d'estimer des paramètres spécifiques à chaque tranche d'âge afin d'étudier cette question. Une étude similaire sur un échantillon plus nombreux permettrait d'explorer de façon plus précise les mécanismes de diffusion du progrès technique par tranche d'âge. Un échantillon plus fourni permettrait également de développer l'inférence relative aux effets mis en évidence dans l'étude. Lors d'analyses par micro-simulations, l'inférence statistique est délicate[19] et la complexité des expressions ne rend l'inférence possible que par rééchantillonnage.

[19]Bourguignon, Ferreira, et Leite (2002) par exemple proposent des décompositions d'indices d'inégalité par micro-simulation, d'où l'inférence est absente.

Chapitre 3

Analyse des inégalités de consommation de soins : l'apport des micro-simulations

3.1 Introduction

La question des inégalités de santé et de recours aux soins apparaît régulièrement sur le devant de la scène. Dans les pays européens, les politiques de santé se donnent en effet pour but non seulement d'améliorer l'état de santé global des populations, mais également de garantir une certaine égalité d'accès aux soins. Dans la plupart de ces pays, une conception égalitaire du système de soins domine : le montant de soins accordé dépend du *besoin*[1] des individus, et est découplé de leur capacité à payer pour ces soins. Le système français de Sécurité Sociale et le *National Health Service* britannique ont un fonctionnement très différent, mais ils ont en commun ces principes de base. En revanche, aux Etats-Unis, malgré la présence de *Medicare* et *Medicaid*, l'égalité d'accès aux soins est une préoccupation moins présente dans l'organisation du système de santé. Pour Wagstaff et van Doorslaer (2000), la littérature portant sur l'égalité et l'équité de recours aux soins est largement liée aux préoccupations politiques, dont témoigne l'abondance de la littérature européenne à ce sujet. La littérature américaine ne s'est ainsi intéressée à ce thème qu'à partir de l'époque des tentatives de réforme de l'administration Clinton, visant à améliorer

[1] Une discussion de la notion de *besoin* de soins est présente dans Wagstaff et van Doorslaer (2000).

l'équité du système de santé américain (Wagstaff et van Doorslaer 2000). Pour ces auteurs, l'équité du système de soins est reconnu comme un objectif fondamental, plus important même que son efficacité. Pour la France, un extrait des ordonnances du 4 octobre 1945 à l'origine de la Sécurité Sociale expose les motifs suivants[2] :

> La sécurité sociale est la garantie donnée à chacun qu'en toutes circonstances il disposera des moyens nécessaires pour assurer sa subsistance et celle de sa famille dans des conditions décentes. Trouvant sa justification dans un souci élémentaire de justice sociale, elle répond à la préoccupation de débarrasser les travailleurs de l'incertitude du lendemain, de cette incertitude constante qui crée chez eux un sentiment d'infériorité et qui est la base réelle et profonde de la distinction des classes entre les possédants sûrs d'eux-mêmes et de leur avenir et les travailleurs sur qui pèse, à tout moment, la menace de la misère.
>
> Envisagée sous cet angle, la sécurité sociale appelle l'aménagement d'une vaste organisation nationale d'entraide obligatoire qui ne peut atteindre sa pleine efficacité que si elle présente un caractère de très grande généralité à la fois quant aux personnes qu'elle englobe et quant aux risques qu'elle couvre. Le but final à atteindre est la réalisation d'un plan qui couvre l'ensemble de la population du pays contre l'ensemble des facteurs d'insécurité ; un tel résultat ne s'obtiendra qu'au prix de longues années d'efforts persévérants, mais ce qu'il est possible de faire aujourd'hui, c'est d'organiser le cadre dans lequel se réalisera progressivement ce plan.

Alors que le système de santé français est réputé égalitaire, avec un régime général couvrant 99% de la population et des tarifs presque tous réglementés (médecins du secteur 1 à tarifs réglementés, prix des médicaments négociés entre les industriels et les Caisses d'Assurance Maladie etc.), des inégalités de santé persistent, et ce au détriment des personnes les plus modestes (Jusot 2003, van Doorslaer et Masseria 2004, Couffinhal, Dourgnon, et Tubeuf 2004). Une explication à cette situation pourrait être la simple corrélation entre pénibilité du travail et bas salaire (Cribier 1997). Néanmoins, si l'on estime que les soins médicaux consommés ont une influence sur l'état de santé individuel, il est logique de s'interroger sur l'accès réel aux soins, et en particulier sur son caractère équitable ou non. Le Régime Général ne couvre en effet pas toute la population. Et même pour les personnes couvertes par le Régime Général, des barrières, en particulier financières, subsistent : avance de frais, dépassements, absence de complémentaire... C'est la raison pour laquelle a été institué le 1er janvier 2000 le dispositif de la Couverture Maladie Universelle (CMU). Depuis cette date, les personnes non couvertes par un quelconque régime obligatoire peuvent demander leur affiliation. En outre, le système de CMU complémentaire

[2]Source : le portail de la Sécurité Sociale `www.securite-sociale.fr`.

(CMU-C) permet aux ménages les plus pauvres de bénéficier d'une couverture complémentaire gratuite (sous condition de ressources) leur permettant de bénéficier du système de santé sans dépense à charge et sans avance de frais[3]. Le plafond de ressources est néanmoins légèrement inférieur au minimum vieillesse, ce qui exclut de fait les personnes âgées de plus de 65 ans.

Si d'aventure toutes les barrières financières au recours aux soins étaient levées, il n'est pas certain que l'on constaterait une distribution de recours aux soins parfaitement équitable. Le caractère équitable de la délivrance des soins de santé peut se définir par deux dimensions : l'équité horizontale et l'équité verticale (Wagstaff et van Doorslaer 2000). L'équité horizontale se définit par la situation où en moyenne, les personnes ayant un même besoin de traitement reçoivent un traitement similaire, indépendamment de leurs autres caractéristiques (en particulier le revenu). L'équité verticale se définit par la situation où des personnes ayant des besoins de soins différents reçoivent un traitement différencié reflétant leurs différences de besoins. Dans le domaine de la recherche sur l'équité de la consommation de soins, l'équité horizontale est la notion la plus souvent étudiée, alors que l'équité verticale est un thème de recherche largement délaissé[4].

Il existe une foule de facteurs pouvant rendre la distribution de consommation de soins inéquitable en apparence. On pense en particulier aux différences de préférences entre les agents : certains accordent plus d'importance à la santé que d'autres par exemple. Il peut également y avoir divergence sur le gain apporté par les soins (ainsi que sur l'évaluation de ce gain) : si le handicap provoqué par la maladie empêche de travailler correctement, le manque à gagner pourra être plus important pour un individu à salaire élevé. Au regard des coûts de soins engagés (pouvant se réduire à des coûts de déplacement ou d'opportunité), cet individu pourra donc être plus incité à se soigner qu'un individu à faible salaire. L'hétérogénéité des fonctions d'utilité d'une part, et l'arbitrage coût/avantage d'autre part peuvent expliquer l'inégalité de consommation de soins observée en fonction du revenu, à

[3]Source : le portail de la CMU `www.cmu.fr`.

[4]Voir Wagstaff et van Doorslaer (2000). Dans le domaine de la recherche portant sur l'équité de *financement* des dépenses de santé, la notion d'équité verticale est beaucoup plus présente : dans nombre de systèmes de santé, les individus contribuent selon leur capacité à payer et la progressivité des contributions est un thème largement traité par la littérature.

niveau de santé donné. L'inégalité résulterait dans ce cas de l'expression des préférences par des choix rationnels.

Quel peut être le but d'un régulateur bienveillant ? Avoir pour but normatif l'égal recours aux soins (équité horizontale) implique de ne pas tenir compte des préférences des agents. Plus simplement, on peut considérer que le régulateur se donne pour objectif un égal accès aux soins. Or l'égalité d'*accès* est difficilement observable en pratique. Se concentrer sur l'égalité de *recours* aux soins à pathologie donnée est alors un moyen d'atteindre l'égalité d'accès. Cet objectif repose sur l'hypothèse que les préférences des individus ont une influence négligeable : dans ce cas, les inégalités de recours sont une bonne mesure des inégalités d'accès. Atteindre l'égalité de santé est parfois conçu comme le but sous-tendant l'égalité d'accès aux soins : les personnes les plus pauvres étant souvent également les plus malades (Jusot 2003), une couverture universelle des soins telle que mise en place en France a pour objectif de gommer les inégalités de recours, et par contrecoup peut espérer gommer les inégalités de santé. Or Wagstaff et van Doorslaer (2000) montrent que l'égalité d'accès aux soins, l'allocation selon le besoin et l'égalité de santé sont des buts non nécessairement compatibles. Ces auteurs, dont les travaux constituent une référence dans la littérature sur le sujet, adoptent le point de vue suivant : une distribution équitable de soins doit refléter le "besoin" de soins dans la population. Toute déviation par rapport à cette distribution (appelée *need-expected*) est considérée comme inéquitable. L'hypothèse sous-jacente est que l'état de santé, même inégalitaire, est une donnée : seule l'inégalité de recours à état de santé donné est analysé[5].

Comme nous l'avons signalé plus haut, à "besoin" (ou morbidité) égal, deux individus peuvent choisir des niveaux de recours aux soins différents. Il s'agit là d'un choix reflétant les préférences de l'individu, préférences qu'il est d'usage de respecter dans le cadre de l'analyse économique de la demande d'un bien. Les soins de santé ne peuvent pas cependant être considérés comme des biens de consommation classiques. En effet, on ne consomme pas des soins de santé pour l'utilité qu'ils procurent intrinsèquement : on

[5]Il faut cependant garder à l'esprit que les inégalités d'état de santé entre individus sont susceptibles d'évoluer et d'être influencées à long terme par les soins reçus.

les consomme car ils améliorent la santé. Il ne s'agit donc pas précisément d'une demande de biens de santé en tant que telle, mais plutôt d'une demande de santé de la part des individus. La santé faisant partie intégrante du capital humain de l'individu (Grossman 1999) au même titre que son instruction par exemple, la santé de la population entre dans une grande part dans les facteurs de croissance d'un pays. Le régulateur peut donc avoir intérêt à contrarier les préférences des individus pour leur assurer un niveau minimum de santé. Le régulateur peut ainsi imposer certaines mesures, afin de tenir compte de la présence d'externalités dans ce domaine (exemple : la vaccination obligatoire). C'est le cas pour l'assurance automobile, obligatoire, alors même que certaines personnes pourraient préférer ne pas s'assurer. C'est également le cas pour l'instruction (scolarité obligatoire jusqu'à 16 ans en France). Parallèlement, l'instruction obligatoire permet de constituer un capital humain individuel "minimum", alors que certaines personnes pourraient ne pas souhaiter investir dans l'éducation. Pour Piketty (2004), "l'enseignement élémentaire obligatoire est sans doute la redistribution efficace[6] la plus importante qui soit, et les travaux sur la croissance et la convergence suggèrent que la progression considérable des niveaux de vie qu'ont connue les pays occidentaux depuis le XIXe siècle ne se serait pas produite sans ces politiques". Cette consommation imposée à tous dans les mêmes proportions, alors qu'elle peut contrarier les préférences, peut être bénéfique au niveau collectif en raison de la présence - entre autres - d'externalités, positives ou négatives.

Outre les mesures imposées, le régulateur peut avoir intérêt à influencer les préférences pour promouvoir l'égalité d'accès aux soins (Wagstaff et van Doorslaer 2000) tout en laissant libre choix de décision aux individus. Les campagnes d'information visent dans ce cadre à permettre aux personnes de prendre leur décision en toute connaissance de cause (campagnes d'informations sur le dépistage par exemple). Le but ultime de ces campagnes d'information, liées à une levée des barrières financières, est bien d'aboutir à une consommation de soins équitable, c'est-à-dire reflétant le besoin dit "objectif" de soins (exemple :

[6]Pour l'auteur, une redistribution efficace corrige des imperfections de marché ; une telle redistribution "améliore l'efficacité parétienne de l'allocation des ressources et l'équité de leur distribution". Elle est opposée à la redistribution "pure" où "des considérations de pure justice sociale exigent une redistribution des individus les mieux dotés vers ceux qui le sont moins", alors que l'équilibre de marché est optimal au sens de Pareto.

pathologies identifiées de manière objective).

Si la consommation de soins reflétait uniquement le besoin de soins au niveau individuel, le niveau de cette consommation devrait être indépendante du revenu. Or au niveau macro-économique, l'élasticité-revenu de la consommation de soins de santé apparaît supérieure à 1 pour les pays développés (Gerdtham et Jönsson 2000) : la consommation de soins aurait les caractéristiques d'un bien de luxe. Les soins de santé étant un bien de confiance (Bien 2001), l'information sur la qualité du service rendu peut passer par la réputation et par le prix, qui constitue souvent un signal précieux. Le signal par les prix est cependant limité par le fait que les tarifs sont fixés en secteur I, qui regroupe 85% des généralistes et 65% des spécialistes[7]. Lorsque chez un spécialiste, le prix des consultations n'est pas assez élevé, la demande peut être faible pour une certaine catégorie de la population en quête de qualité et prête à débourser un prix élevé. La hiérarchie existant entre les tarifs des médecins spécialistes et généralistes (respectivement 25 Euros et 21 Euros en secteur I) peut être identifiée par les consommateurs comme une différence de qualité. Les séances de médecin généraliste sont ainsi susceptibles d'apparaître comme un bien inférieur, et les séances de spécialiste comme un bien supérieur. La politique récente d'introduction de *gatekeeper* dans le système de soins français [8] peut être vue comme une tentative du régulateur de sortir la consommation de soins de généralistes de son statut de bien inférieur.

Nous verrons dans la suite que la consommation de soins de généralistes est concentrée chez les plus pauvres, et que la consommation de soins de spécialistes est concentrée chez les plus riches, ce que l'on retrouve dans la littérature (van Doorslaer et Masseria 2004). Cependant, le lien entre la nature du bien considéré (inférieur ou supérieur) et son indice de concentration dans la population en fonction du revenu n'est pas évident. La qualité de bien inférieur des séances de généralistes n'impliquerait pas la concentration de la consommation de ce bien chez les plus pauvres. Prenons par exemple une population ayant pour revenu total R, et se réduisant à 2 individus, l'un "pauvre" et l'autre "riche". Le

[7]Source : le portail de l'Assurance Maladie `www.ameli.fr`

[8]Depuis 2006, obligation de choix d'un médecin référent, puis obligation de passer par lui pour être remboursé au tarif normal d'une visite chez tout autre médecin

"pauvre" dispose du revenu $0,2R$, et le "riche" du revenu $0,8R$. Intéressons-nous à un bien inférieur : la propension à consommer ce bien est donc plus élevée chez la personne pauvre que chez la personne riche. Soient respectivement $0,3$ et $0,2$ ces propensions, observées chez le pauvre et le riche. La consommation de ce bien sera donc de $0,3 * 0,2R = 0,06R$ pour le pauvre et $0,2 * 0,8R = 0,16R$ pour le riche. Alors même que le bien en question est un bien inférieur, sa consommation est concentrée chez la personne riche.

Ce chapitre propose une méthode originale de mesure des inégalités de consommation de soins, dans un cadre permettant l'évaluation de l'impact des préférences individuelles sur les inégalités de recours aux soins. Il est aussi possible, dans une certaine mesure, d'identifier les composantes de ces préférences du point de vue de l'offre et de la demande de soins.

Nous exposons tout d'abord les méthodes couramment employées par la littérature sur les inégalités de recours aux soins, et la décomposition par facteur que l'on y trouve habituellement. Puis, nous exposons la méthodologie alternative proposée pour décomposer l'inégalité observée par facteurs. Enfin, nous procédons à trois applications de cette méthode sur données françaises :

- la décomposition par facteur des inégalités de consommation de séances de médecins en France, pour l'année 1998
- l'identification de l'impact de l'introduction de la CMU-C dans l'évolution des inégalités de consommation de soins observée entre les années 1998 et 2002
- la comparaison des facteurs menant à l'inégalité de consommation de soins entre les personnes de plus de 65 ans et les autres, pour l'année 1998.

La méthodologie originale proposée permet d'identifier des effets non discernables dans les études actuellement disponibles.

Les données proviennent de l'Appariement construit par l'IRDES[9]. Il s'agit de l'enquête SPS (Santé Protection Sociale) menée tous les deux ans sur un sous-échantillon de l'EPAS (Echantillon Permanent d'Assurés Sociaux) de la CNAMTS, pour les années 1998

[9]Institut de Recherche et de Documentation en Economie de la Santé, Paris

et 2002, choisies pour des raisons d'effectif suffisant (l'échantillon enquêté a été en effet doublé dès 1998). Les effectifs sont de plus de 4000 personnes exploitables par enquête. On dispose dans cette enquête des relevés de consommation de chaque individu, ainsi que de ses caractéristiques individuelles comme l'âge et le revenu par unité de consommation, mais également d'indicateurs synthétiques de morbidité (ex : degré d'invalidité) et de l'indication des pathologies dont souffre l'individu. Les niveaux de revenu sont corrigés de l'inflation (base 1992, par cohérence avec le chapitre 2). Toutes les valeurs sont converties en Euros. Une description plus détaillée des données est fournie en section 2.3.

3.2 Décomposition des inégalités de recours aux soins : la méthodologie standard et ses limites

De nombreuses études (van Doorslaer et Masseria 2004, van Doorslaer, Wagstaff, Burg, Christiansen, Graeve, Duchesne, Gerdtham, Gerfin, Geurts, Gross, Hakkinen, John, Klavus, Leu, Nolan, O'Donnell, Propper, Puffer, Schellhorn, Sundberg, et Winkelhake 2000, van Doorslaer, Koolman, et Jones 2004) montrent que le rang dans la distribution des revenus est très corrélé au rang dans la distribution d'état de santé : les "riches" sont proportionnellement en meilleure santé que les "pauvres", dans les pays en développement, mais également dans les pays développés. Il apparaît également que pour certains postes de dépense (en particulier les séances de médecin spécialistes), les "riches" consomment proportionnellement plus de soins de santé que les "pauvres" (ce qui n'est pas le cas pour les séances de généralistes), ce qui peut apparaître comme inéquitable : les besoins seraient concentrés chez les "pauvres" et la consommation de soins concentrée chez les "riches". Ce cas particulier de la demande de soins de spécialistes se retrouve dans beaucoup de pays européens (van Doorslaer et Masseria 2004).

3.2.1 L'indice de concentration

Un indice de mesure de l'inégalité de consommation de soins en fonction du revenu doit être en mesure de décrire l'inégalité de consommation observée, tout en mettant en évidence la part d'inégalité due aux besoins différenciés selon le revenu (part "justi-

fiable" de l'inégalité), et la part d'inégalité due aux autres caractéristiques de l'individu (part "injustifiable"). La distinction entre inégalité de recours aux soins justifiable et injustifiable est d'usage courant dans cette littérature (voir Wagstaff et van Doorslaer (2000) par exemple). L'indice de mesure d'inégalité utilisé ici sera l'indice de concentration des soins en fonction du revenu. Cet indice est largement répandu, et possède de bonnes propriétés (Wagstaff et van Doorslaer 2000). Il est utilisé en particulier par les travaux du groupe ECUITY[10] dans le cadre de comparaisons internationales d'inégalités de recours aux soins.

L'indice de concentration d'une variable y (par exemple le nombre de séances de médecin) en fonction du revenu R se définit à partir de la courbe de concentration de y en fonction de R. Cette courbe de concentration relie le pourcentage cumulé des individus de la population classés par revenu croissant à la proportion de y consommé par ce pourcentage, par rapport à la consommation totale. La mesure de l'inégalité de recours aux soins est donnée dans le cas de l'indice de concentration par deux fois l'aire entre la courbe de concentration et la première bissectrice (droite d'égalité). Soit une population de n individus notés $i \in [1..n]$ ayant chacun un recours au soins y_i et un revenu W_i, avec $\overline{y}$ la moyenne des y_i et R_i le rang relatif *(fractional rank)* de l'individu i dans la distribution des revenus[11]. C, l'indice de concentration de y pour cette population, s'écrit (Wagstaff, van Doorslaer, et Watanabe 2003) :

$$C = \frac{2}{n\overline{y}} \sum_{i=1}^{n} y_i R_i - 1 \tag{3.1}$$

L'indice de concentration de la variable y en fonction de la variable R peut se calculer en estimant la régression suivante WorldBank (2005a) :

$$2\sigma_R^2 \frac{y_i}{\overline{y}} = \alpha + \beta R_i + v_i \tag{3.2}$$

[10]Projet européen visant à étudier les inégalités de santé et l'inéquité de recours aux soins et de financement des soins : `http://www2.eur.nl/bmg/ecuity/`. Les équipes de recherche proviennent des pays suivants : Belgique, Danemark, Finlande, France, Allemagne, Grèce, Irlande, Italie, Pays-Bas, Espagne, Suède, Royaume-Uni, Norvège, Suisse, Etats-Unis.

[11]Soit une population de n individus, et soit r_i le rang de l'individu i dans la distribution des revenus. Le *fractional rank* R_i se définit de la façon suivante : $R_i = \frac{r_i}{n}$.

où R_i est le rang relatif de l'individu i dans la distribution des revenus et σ_R^2 sa variance. L'estimateur $\widehat{\beta}$ est égal à l'indice de concentration de la variable y par rapport au revenu, et l'écart-type de $\widehat{\beta}$ donne également l'écart-type de l'indice de concentration. L'estimation est menée par une méthode de Newey et West (1994) afin de corriger de l'éventuelle autocorrélation ou hétéroscédasticité pouvant être présente dans les données.

L'indice de concentration en tant que résumé synthétique du degré d'inégalité de la consommation de soins parmi la population souffre un certain nombre de limites. Ces limites sont analogues à celles de l'indice de Gini, dont l'indice de concentration est dérivé. L'indice de concentration dépend du rang des individus dans la distribution des revenus, mais non de leur niveau de revenu. D'autre part, cet indice accorde le même poids à l'inégalité parmi les riches qu'à l'inégalité parmi les pauvres (Gourieroux 1999). On ne s'attend donc pas à ce que cet indice répercute de manière très sensible les augmentations de consommation de soins chez les plus pauvres. De plus, l'indice de concentration ne peut pas fournir un ordre complet permettant de classer toutes les distributions en fonction de leur degré d'inégalité. Soit une distribution $(x_i), 1 \leq i \leq n$ dont l'indice de concentration est inférieur à celui d'une autre distribution $(y_i), 1 \leq i \leq m$: (x_i) ne domine pas forcément (y_i) du point de vue de l'inégalité. En effet, les deux courbes de concentration (à partir desquelles sont définis les indices de concentration) peuvent se croiser. Dans ce cas, aucune des deux courbes n'étant constamment sous l'autre, une distribution admettra une égalité plus grande dans les hauts revenus, alors que l'autre admettra une égalité plus grande dans les bas revenus. De même, deux distributions ayant le même indice de concentration n'ont aucune raison d'être égales, et une distribution ayant pour indice de concentration 0 n'est pas forcément parfaitement égalitaire. La courbe de concentration peut croiser la droite d'égalité : une partie de la courbe peut se trouver au dessus de la droite d'égalité, et l'autre partie dessous, contrairement à ce que l'on a pour les courbes de Lorenz classiques décrivant la répartition du revenu. Pour toutes ces raisons, l'indice de concentration est un outil imparfait, mais nous le conserverons pour un objectif de comparabilité de nos résultats avec la littérature préexistante. Soulignons néanmoins que notre analyse est fondée sur des distributions simulées dont on mesure le degré d'inégalité : il est tout à fait possible de

mesurer le degré d'inégalité de chacune de ces distributions d'une autre façon qu'avec l'indice de concentration.

3.2.2 La décomposition par facteurs de l'indice de concentration

On suppose ici que le recours aux soins est expliqué par des facteurs $(x_k), k \in [1..K]$ avec un modèle linéaire standard (Wagstaff, van Doorslaer, et Watanabe 2003). Une hypothèse importante de la méthode est que la consommation de soins peut être correctement décrite par une spécification linéaire.

$$\forall i \in [1..n], y_i = \alpha + \sum_{k=1}^{K} \beta_k x_{k,i} + \varepsilon_i \tag{3.3}$$

Ce qui implique la décomposition suivante (Wagstaff, van Doorslaer, et Watanabe 2003)[12] :

$$C = \sum_{k=1}^{K} \frac{\beta_k \overline{x_k}}{\overline{y}} C_k + \frac{GC_\varepsilon}{\overline{y}} \tag{3.4}$$

Où $\overline{x_k}$ est la moyenne des $x_{k,i}$, $\overline{y}$ la moyenne des y_i, C_k l'indice de concentration de x_k par rapport au revenu, et GC_ε l'indice de concentration généralisé de ε_i ($GC_\varepsilon = \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i R_i$).

Pour étudier les changements dans le degré d'inégalité entre deux périodes données (ΔC), Wagstaff, van Doorslaer, et Watanabe (2003) proposent la décomposition suivante, directement inspirée de la décomposition linéaire (3.4) :

$$\Delta C \approx dC = -\frac{C}{\overline{y}} d\alpha + \sum_{k=1}^{K} \frac{\overline{x_k}}{\overline{y}} (C_k - C) d\beta_k + \sum_{k=1}^{K} \frac{\beta_k}{\overline{y}} (C_k - C) d\overline{x_k} + \sum_{k=1}^{K} \frac{\beta_k \overline{x_k}}{\overline{y}} dC_k + d\frac{GC_\varepsilon}{\overline{y}} \tag{3.5}$$

Les variations observées de l'indice de concentration du recours aux soins sont fonction des variations des indices de concentration de chacun des facteurs expliquant ce recours.

[12] Wagstaff, van Doorslaer, et Watanabe (2003) développent cette décomposition pour dans le cadre de l'analyse de l'inégalité de *santé* et non de *recours aux soins*. Cette décomposition est cependant applicable à toute variable y (voir van Doorslaer et Masseria (2004)).

3.2.3 Une définition de l'inéquité : le Horizontal Inequity Index

L'inégalité de consommation de soins en fonction du revenu peut s'expliquer entièrement par des besoins différenciés selon le revenu : si la distribution de la consommation de soins rend entièrement compte de la distribution des besoins, la distribution peut ainsi être inégalitaire mais parfaitement équitable. Afin d'analyser plus précisément l'inéquité des soins, la littérature adopte l'indice d'inéquité horizontale (*Horizontal Inequity Index* par Wagstaff et van Doorslaer (2000)). L'inéquité de la consommation de soins est définie comme la part d'inégalité de consommation qui n'est pas expliquée par les besoins : il s'agit de la part d'inégalité "injustifiable". Reprenant l'expression (3.4), et notant M les variables dénotant le "besoin" et Z les autres, et baptisant respectivement δ et γ leurs vecteurs de coefficient associés, on a :

$$C = \sum_{k=1}^{K} \frac{\delta_k \overline{m_k}}{\mu} C_k + \sum_{k=1}^{K} \frac{\gamma_k \overline{z_k}}{\mu} C_k + \frac{GC_\varepsilon}{\mu}. \tag{3.6}$$

Le premier terme de l'expression (3.6) représente l'inégalité "justifiable", car provoquée par l'inégalité des besoins (le *Need*). Le second terme représente l'inégalité "non justifiable", car provenant de facteurs autres que le besoin de soins : il s'agit du degré d'inéquité horizontale, noté *HI* (*Horizontal Inequity Index*). On a donc, en conservant ces notations :

$$Need = \sum_{k=1}^{K} \frac{\delta_k \overline{m_k}}{\mu} C_k \tag{3.7}$$

et

$$HI = \sum_{k=1}^{K} \frac{\gamma_k \overline{z_k}}{\mu} C_k + \frac{GC_\varepsilon}{\mu}. \tag{3.8}$$

HI représentant l'indice d'inéquité horizontale, il est analysé en la somme de ces composantes. L'impact des variables hors morbidité est ainsi analysé.

3.2.4 Une décomposition nécessairement linéaire

Selon la méthode de décomposition (3.4), la décomposition de l'indice de concentration de y est fonction de la somme des indices de concentration des facteurs explicatifs

x_k, chaque C_k étant pondéré par l'élasticité de y par rapport à x_k, évaluée à l'individu moyen. Cette somme comporte également un terme résiduel englobant l'inégalité non expliquée par le modèle. Lors d'une estimation de cette décomposition, les β sont remplacés par les $\hat{\beta}$ estimés et le terme résiduel est calculé avec les résidus estimés $\hat{\varepsilon}$.

L'inconvénient évident de cette décomposition est qu'elle est linéaire. Or, il existe de nombreux modèles explicatifs non linéaires en économie de la santé (voir chapitre 1). Pour expliquer un nombre de consultations, on a par exemple un modèle de comptage de type *GLM* de la forme :

$$E(y_i) = h(\alpha + \sum_{k=1}^{K} \beta_k x_{k,i}) \tag{3.9}$$

Où h est une fonction non linéaire (souvent la fonction exponentielle) et où l'on suppose que la variable y suit une loi de Poisson ou une loi Gamma. Les paramètres β_k sont alors estimés par maximum de vraisemblance (voir section 1.2.3 pour une description de l'estimation).

Dans ce cas, la forme non linéaire de l'expression 3.9 ne permet pas d'évaluer distinctement les effets des différents facteurs x_k. Force est d'avoir recours à une linéarisation de 3.9. On utilise alors la décomposition suivante :

$$C = \sum_{k=1}^{K} \frac{\beta_k^m \overline{x_k}}{\mu} C_k + \frac{GC_\varepsilon}{\bar{y}} \tag{3.10}$$

où les β_k^m sont les *effets marginaux* évalués à la moyenne (WorldBank 2005b). La décomposition est donc approximative. Dans le cadre de la décomposition de l'inégalité par facteur, van Doorslaer et Masseria (2004) préconisent d'utiliser un modèle linéaire tel (3.3) pour la variable y, même lorsque ce modèle estimé est inapproprié à la variable en question (par exemple, un nombre de séances, nécessairement positif ou nul). Cette solution demeure toutefois conceptuellement insatisfaisante.

3.2.5 L'interprétation des C_k

La présence dans les formules (3.4), (3.5) et (3.10) d'indices de concentration de chacun des facteurs impose que cet indice ait effectivement un sens. Des indicateurs d'état de santé sont en effet présents dans le modèle explicatif de recours aux soins (3.3 ou 3.9). Les décompositions standard nécessitent donc de calculer et d'interpréter les indices de concentration de ces indicateurs d'état de santé.

Prenons l'exemple d'une mesure d'état de santé comme le *HUI (Health Utility Index)*. Cette mesure possède des propriétés qui permettent d'interpréter son indice de concentration : cet indice rend bien compte de la répartition du "stock" de santé dans la population. Le *HUI* est en effet construit pour chaque individu comme un score d'utilité correspondant à son état de santé, correspondant à une valeur située entre 0 (mort) et 1 (pleine santé), un score pouvant exceptionnellement être négatif si l'état de santé correspondant est jugé pire que la mort. Lors d'une enquête en population générale, on demande aux individus enquêtés leur préférence relative pour un état de santé particulier, par rapport à la pleine santé. Cet état de santé est caractérisé par un certain nombre d'*attributs* $(A_k)_{k=1..K}$ codés de manière catégorielle (niveau de douleur, degré de mobilité ...). L'enquête permet d'évaluer au niveau de l'échantillon le score d'utilité associé à chaque degré $A_{i,k}$ de chacun de ces attributs A_k. Une fonction d'utilité multi-attribut est ensuite utilisée pour associer un *score* d'utilité à tout état de santé E caractérisé par ces mêmes attributs. Une telle fonction est par exemple pour *HUI* 2 (1996)[13] :

$$U(E) = U(A_1, A_2, ..., A_K) = 1,06 * \prod_{k=1}^{K} A_k - 0,06. \tag{3.11}$$

L'hypothèse est que tout état de santé E peut être convenablement décrit par un vecteur $(A_1 \ ... \ A_K)$ et résumé par le score $U(E)$. Pour *HUI* 2, le nombre de combinaisons possibles est de 24 000 états de santé différents. Soulignons que les scores liés à chaque degré de chaque attribut sont évalués grâce à une enquête unique réalisée sur population canadienne. Selon ses auteurs, ces scores sont applicables à l'évaluation de l'état de santé

[13]Source : portail du Health Utility Group de McMaster `http://www.fhs.mcmaster.ca/hug/`

de n'importe quel individu de la planète.

Outre le fait que l'universalité du *HUI* peut être source de débat, les questions nécessaires à son calcul ne sont pas systématiquement présentes dans toutes les enquêtes. Celle dont nous disposons (l'Appariement de l'IRDES) ne les propose pas. L'Appariement propose en revanche des indicateurs d'état de santé par pathologie (présence ou absence de pathologies identifiées, telle le diabète par exemple), ainsi que des indicateurs d'état de santé synthétiques catégoriels, c'est-à-dire prenant leurs valeurs dans N. Par exemple, l'indice de risque vital varie de 0 à 5, et l'indice d'invalidité varie de 0 à 6 (voir annexe B.1.3). Utiliser la décomposition (3.4) en employant des variables explicatives catégorielles mène à devoir interpréter leur C_k. Or l'indice de concentration d'une mesure d'état de santé catégorielle comme le degré d'invalidité est difficilement interprétable. Si l'indice de concentration est calculé sur la variable brute, cela suppose que les degrés d'invalidité peuvent s'additionner (un individu avec un degré d'invalidité de 2 serait 2 fois plus invalide qu'un individu ayant un degré d'invalidité de 1). Pour éviter une telle hypothèse, on peut calculer un indice de concentration pour chacune des six modalités du degré d'invalidité (codées en 0 ou 1). Quoique informatifs, ces indices (qui seraient aussi nombreux que le nombre de modalités de chaque variable catégorielle) rend l'évaluation de la répartition de la santé difficilement interprétable, et en aucun cas synthétique. C'est cette option qui est utilisée par le groupe ECUITY. Une dernière solution fait appel à une transformation afin d'analyser non pas le degré d'invalidité tel quel mais plutôt un "score de santé" ayant de bonnes propriétés. Par exemple, Perronnin, Rochaix, et Tubeuf (2006) proposent la construction d'un indicateur de santé synthétique unidimensionnel agrégeant degré d'invalidité et degré de risque vital. La question des indicateurs de santé synthétiques constitue un champ de recherche à part entière, situé en amont des travaux sur les inégalités de consommation de soins, et nécessaires à l'interprétation de l'expression (3.4) utilisée dans l'approche standard. La méthode que nous proposons plus loin permet d'éviter la définition d'un indicateur de santé synthétique et d'éviter le calcul d'indices de concentration sur des variables dichotomiques.

Par ailleurs, le fait que le degré d'inégalité s'exprime en fonction des indices de concentration de chaque variable implique donc de manipuler l'indice de concentration du *log* du revenu lorsque cette variable est introduite dans l'analyse (van Doorslaer, Koolman, et Jones 2004). Introduire le *log* du revenu dans l'estimation permet d'obtenir des résidus ayant de bonnes propriétés. La transformation *log* conserve l'ordre de classement des individus dans la distribution des revenus *mais* l'indice de concentration de cette variable sera plus faible que si l'on utilisait le revenu non transformé. L'interprétation de l'indice de concentration du *log* du revenu n'est d'ailleurs pas immédiate.

3.2.6 La difficile prise en compte des préférences

L'hypothèse sous-jacente à la décomposition de l'inégalité entre part justifiable et part non justifiable (cf l'expression (3.6)) est la suivante : la distribution équitable de soins est celle que l'on observerait si tous les individus étaient soignés comme l'individu moyen de l'échantillon. En effet, l'estimation du modèle (3.3) réalisée en vue de la décomposition (3.4) puis de la décomposition (3.6) est réalisée sur tout l'échantillon : les paramètres δ_k relatifs aux variables de besoin M reflètent la relation *moyenne* existant entre soin et besoin de soins, sur tout l'échantillon. Toute déviation par rapport à cette relation *moyenne* est prise en compte par les variables Z et leurs coefficients γ. On ne peut donc distinguer des δ_k par niveau de revenu, et des préférences éventuellement différentes selon le niveau de revenu des individus. La décomposition (3.6) n'est en effet faite qu'à partir de la distribution des variables explicatives selon le revenu, et ne prend pas en compte l'éventuelle hétérogénéité des coefficients selon le revenu. Pour tout indicateur de besoin de soins Z_k, le coefficient $\hat{\delta_k}$ estimé sur tout l'échantillon représente ce que la société alloue *en moyenne* comme montant de soins pour ce niveau de besoin. Dans les analyse d'ECUITY, ce comportement moyen est constitué en norme dont la pertinence est non discutée.

3.2.7 La limitation à une équation unique expliquant la consommation de soins

Le chapitre 1 a montré que les modèles explicatifs de la dépense de soins pouvaient être hautement non linéaires. En raison des caractéristiques particulières des dépenses de santé observées au niveau individuel, les spécifications les plus adéquates sont des modèles

en deux équations, expliquant d'une part la participation au système de soin, d'autre part le niveau de consommation conditionnel à cette participation. Les deux équations sont estimées de façon indépendantes, la première grâce à un *logit* ou un *probit*, et la seconde grâce à un *GLM* impliquant également des fonctions non linéaires des variables explicatives (voir section 1.3). La consommation de soins s'écrit donc comme le produit de deux expressions non linéaires : probabilité de participation multipliée par le niveau de consommation conditionnelle. Soulignons qu'une autre façon de modéliser la consommation de soins est la famille de modèles *zero-inflated* à une seule équation. Ces modèles ont cependant des caractéristiques moins proches des données (voir section 1.4).

Or, que ce soit dans un cas d'un modèle explicatif linéaire (modèle (3.3)) ou non linéaire (modèle (3.9)), la décomposition de l'inégalité par facteurs (expressions respectivement (3.4) et (3.10)) n'existe que pour une équation unique expliquant la dépense de soins. Ainsi, les décompositions prenant en considération les caractéristiques particulières des dépenses de santé ne peuvent-elles décomposer qu'une seule équation à la fois. Dans les travaux d'Ecuity sont donc proposées séparément (Wagstaff, van Doorslaer, et Watanabe 2003) :

- une décomposition de la probabilité de participation (estimation par modèle *logit* ou *probit*)
- une décomposition de la consommation conditionnelle à la participation (estimation par modèle *GLM*, par *MCO*...)
- une décomposition de la consommation totale (estimation par modèle *zero-inflated*).

Outre l'approximation linéaire nécessaire à la décomposition dans le cadre de modèles non linéaires, la question se pose de la cohérence entre les trois décompositions. En effet, on a montré dans la section 1.3 que les modèles en deux équations et les modèles *zero-inflated* étaient des modèles conceptuellement différents : la cohérence entre les décompositions précitées n'est pas garantie. Par ailleurs, il est difficile d'effectuer à partir de plusieurs décompositions (une pour la participation, l'autre pour la consommation conditionnelle) une évaluation synthétique de l'impact des différents facteurs d'inégalité.

3.3 Une autre méthodologie par micro-simulations

La décomposition des inégalités de consommation de soins proposée ici permet de surmonter les limites de l'approche traditionnelle, évoquées en section 3.2. La mesure employée est toujours l'indice de concentration, par souci de comparabilité avec les résultats antérieurs, en particulier ceux du groupe ECUITY. Nous nous basons sur des distributions de consommation de soins simulées sous différentes hypothèses. Pour chacune de ces distributions simulées, nous calculons l'indice de concentration. Cette méthode originale offre plusieurs avantages : elle s'affranchit de l'obligation d'approximation linéaire que la décomposition traditionnelle impose lorsque le modèle explicatif de recours aux soins est non-linéaire ; elle évite de se limiter à une seule équation lors de la décomposition. Enfin, elle permet d'évaluer l'impact de l'hétérogénéité des comportements sur les inégalités.

Alors que la décomposition traditionnelle donne l'indice de concentration en fonction des indices de concentration de chacun des facteurs du modèle explicatif (voir les expressions (3.4) et (3.10)), la méthode proposée ici n'impose pas l'utilisation de facteurs explicatifs du recours aux soins dont l'indice de concentration ait un sens.

Prendre des distributions simulées dont on calcule l'indice de concentration *au lieu de* décomposer l'indice de concentration d'une distribution observée permet de mettre en évidence séparément les effets de la participation et de la consommation conditionnelle. En effet, avec la décomposition standard, il est nécessaire de décomposer séparément l'indice de concentration de la participation, de la dépense conditionnelle, et du total. Les résultats de la décomposition peuvent ne pas être cohérents, car les modèles estimés sont conceptuellement différents. Ici, les distributions sont simulées grâce à des modèles en deux parties. Les distributions étant simulées sous différentes hypothèses, les effets relatifs à la participation et à la consommation conditionnelle peuvent être évalués dans le cadre d'un seul et même modèle expliquant la consommation totale de soins.

Enfin, la caractéristique la plus novatrice de cette autre méthode de décomposition est rendre possible une estimation d'un modèle explicatif de la dépense de soins par

quantile de revenu. En effet, un modèle estimé sur tout l'échantillon seulement ne permet pas de distinguer les préférences des individus, pouvant varier selon le quantile de revenu auquel ils appartiennent. Dans un modèle du type (3.3) ou (3.9), pouvoir estimer un β par quantile de revenu permet de repérer les différences dans les préférences des individus. Dans l'expression (3.6), la décomposition du degré d'inéquité horizontale *HI* ne dépend que de la distribution des variables Z par rapport au revenu, et des paramètre γ estimés sur tout l'échantillon. La méthode que nous proposons ici permet de calculer un indice d'inéquité horizontale composé d'une partie due à la distribution des variables Z selon le revenu *et* d'une partie due à l'hétérogénéité des paramètre estimés γ selon le revenu. On pense ainsi fournir un moyen d'identifier les préférences des individus, ce qui n'est pas possible dans la décomposition traditionnelle décrite en section 3.2.

3.3.1 Le choix d'un modèle explicatif du recours aux soins

Soit la variable d'intérêt y_i matérialisant le recours aux soins de l'individu i, appartenant à une population de N individus. Nous allons nous intéresser au nombre de séances chez les médecins généralistes, au nombre de séances chez les médecins spécialistes, ainsi qu'au nombre total de séances. Suivant les résultats du chapitre 1, le choix ici est de prendre une spécification en deux équations indépendantes (*modified Two Part Model*), avec en première partie une spécification Logit, et en deuxième partie une spécification *GLM*. La spécification choisie est la loi Binomiale Négative tronquée à 0 avec lien *log* pour modéliser le nombre de séances chez les praticiens (on retrouve les spécifications couramment employées dans la littérature). Le modèle explicatif s'écrit donc :

(1) $P_i = I_{P_i^*>0}$ et $E(P_i^*) = G(Z_1\gamma_1)$ avec $G(p) = \dfrac{1}{1+e^{-p}}$

(2) $y_i = I_{P_i=1}y_i^*$ et $E(y_i^*/P_i = 1) = F(Z_2\gamma_2)$ avec $F(p) = e^p$ et y_i suivant la loi Binomiale Négative tronquée à 0.

Par conséquent, $E(y_i) = E(P_i^*) * E(y_i^*/P_i = 1)$

3.3.2 Une décomposition de l'indice de concentration du recours aux soins

Soit C l'indice de concentration du recours aux soins observé sur un échantillon donné. Il s'agit de l'indice de concentration de la variable y par rapport au revenu R. Le recours aux soins étant expliqué par un certain nombre de facteurs explicatifs (les vecteurs de variables Z_1 et Z_2), il est essentiel d'identifier la contribution de chacun de ces facteurs à la valeur observée de C. En effet, les inégalités de recours aux soins en fonction du revenu peuvent provenir par exemple de facteurs "justifiables" (un besoin différent selon le quantile de revenu auquel on appartient) ou bien "non-justifiables" (une discrimination dans l'accès aux soins) (Wagstaff et van Doorslaer 2000).

La démarche exposée ici vise de manière très générale à identifier l'apport de certains groupes de variables au degré d'inégalité observé. On a vu que la littérature distingue les variables de "besoin" (morbidité, âge ...) des variables de "contrôle" (catégorie socio-professionnelle, niveau d'études ...). On procèdera donc à l'exposé de la décomposition en distinguant ces 2 groupes de variables, tout en gardant à l'esprit que cette décomposition peut s'appliquer à n'importe quel groupe de variables.

Les vecteurs de variables explicatives Z_1 et Z_2 sont subdivisés en 2 groupes de variables : des variables de besoin notées respectivement M_1 et M_2 ainsi que des variables de contrôle notées respectivement X_1 et X_2. L'objectif est ici de décomposer la valeur de l'indice C en la somme des contributions de chaque groupe de variables, en utilisant des distributions simulées pour le recours aux soins.

Le modèle devient :

(1) $E(P_i) = G(X_1 b + M_1 d)$

(2) $E(y_i/P_i = 1) = F(X_2\beta + M_2\delta)$

Et la prédiction du modèle serait, pour tout individu i :

$$\widehat{y_i} = \underbrace{G(X_{1,i}\widehat{b} + M_{1,i}\widehat{d})}_{participation} * \underbrace{F(X_{2,i}\widehat{\beta} + M_{2,i}\widehat{\delta})}_{consommation\ conditionnelle} \tag{3.12}$$

3.3.3 Estimation des paramètres

Le modèle est estimé sur l'échantillon entier : on obtient donc les vecteurs de paramètres $\widehat{b}$, $\widehat{d}$, $\widehat{\beta}$ et $\widehat{\delta}$. L'échantillon est ensuite divisé en Q quantile de revenus. Le modèle est alors estimé sur chacun de ces Q quantiles. On obtient donc les vecteurs de paramètres $(\widehat{b_q})$, $(\widehat{d_q})$, $(\widehat{\beta_q})$ et $(\widehat{\delta_q}), q \in [1..Q]$. Les vecteurs de paramètres $\widehat{\beta_q}$ peuvent différer entre eux, ce qui indiquerait des pratiques différentes selon le quantile de revenu auquel appartient l'individu. De même, les paramètres $\widehat{\delta_q}$ peuvent différer entre eux, ce qui indiquerait un effet variable sur la consommation de soins des caractéristiques socio-économiques selon le revenu (par exemple, un effet différent de la couverture complémentaire selon le revenu). Cette estimation de coefficients spécifiques à des sous-groupes distincts de population se retrouve dans les études inspirées d'Oaxaca (1973). Dans le cas où l'on cherche à évaluer l'impact sur les inégalités de recours aux soins d'une mesure (notée MES) visant une certaine catégorie de la population de faible effectif (par exemple, la CMU-C), on estime $\widehat{b}$, $\widehat{d}$, $\widehat{\beta}$ et $\widehat{\delta}$ (ainsi que les $\widehat{b_q}$, $\widehat{d_q}$, $\widehat{\beta_q}$ et $\widehat{\delta_q}$) sur les non-bénéficiaires de la mesure uniquement. On estime ensuite le modèle séparément pour les bénéficiaires de la mesure, et on obtient des paramètres estimés $\widehat{b_{MES}}$, $\widehat{d_{MES}}$, $\widehat{\beta_{MES}}$ et $\widehat{\delta_{MES}}$.

Pour chaque équation estimée, les vecteurs de paramètres estimés représentent les comportements individuels. Parmi ces vecteurs, les vecteurs de paramètres b et β (ainsi que leurs équivalents b_q et β_q au niveau des différents quantiles de revenu q) représentent les pratiques observées à morbidité donnée, toutes choses égales par ailleurs. Dans la suite, on parlera de *pratiques* pour désigner ces paramètres : en effet, la valeur d'un paramètre estimé représente la réponse en terme de quantité de soins à un état de santé particulier. Ces *pratiques* relèvent des comportements individuels, et peuvent relever autant de la

demande que de l'offre de soins. Les paramètres b sont relatifs aux pratiques liées à la participation : ces pratiques sont donc principalement liées à la demande. Les paramètres β sont relatifs aux pratiques liées à la consommation conditionnelle : ces pratiques sont donc principalement liées aux comportements des offreurs de soins, et dans une moindre mesure à la demande.

3.3.4 Simulation des distributions et calcul de leurs indices de concentration

Une fois ces paramètres estimés, on procède à des simulations de distribution. Afin de comprendre le rôle de chaque facteur dans l'indice d'inégalité observé, on part d'une distribution de recours aux soins parfaitement égalitaire ($\widehat{y_i^0}$), puis on fait varier un par un chaque facteur explicatif du recours aux soins pour identifier son apport au degré d'inégalité observé sur la variable y_i. Les différentes étapes des distributions simulées sont décrites ci-après (on note $\overline{X_1}$, $\overline{M_1}$, $\overline{X_2}$ et $\overline{M_2}$ les vecteurs de variables correspondant à l'échantillon moyen).

Le point de départ des micro-simulations consiste en une distribution de consommation de soins parfaitement égalitaire. Cette distribution égalitaire correspond à la distribution où tous les individus partageraient les mêmes caractéristiques ($\forall i,\ \forall j \in \{1,2\}, X_{j,i} = \overline{X_j}$ et $M_{j,i} = \overline{M_j}$) et où l'impact de ces caractéristiques serait le même, quel que soit le quantile de revenu auquel appartient l'individu ($\forall q, b_q = b = b_{MES}$ et $d_q = d = d_{MES}$ et $\forall q, \beta_q = \beta = \beta_{MES}$ et $\delta_q = \delta = \delta_{MES}$). Cette distribution du recours aux soins parfaitement égalitaire peut s'écrire :

$$\forall i, \widehat{y_i^0} = G(\overline{X_1}\widehat{b} + \overline{M_1}\widehat{d}) * F(\overline{X_2}\widehat{\beta} + \overline{M_2}\widehat{\delta}) = Constante \tag{3.13}$$

Le recours aux soins ici simulé étant par définition le même pour tous, l'indice de concentration du recours aux soins est mécaniquement égal à 0 : $CI_0 = 0$.

Les étapes successives de simulation sont les suivantes :

1) La distribution permettant de mettre en évidence l'effet sur le recours aux soins de la morbidité (différenciée selon les individus), à comportements constants, peut s'écrire de la façon suivante :

$$\forall i, \widehat{y_i^1} = G(\overline{X_1}\widehat{b} + M_{1,i}\widehat{d}) * F(\overline{X_2}\widehat{\beta} + M_{2,i}\widehat{\delta}) \tag{3.14}$$

Où l'on a remplacé $\overline{M_1}$ par $M_{1,i}$ et $\overline{M_2}$ par $M_{2,i}$ dans l'expression (3.13). L'indice de concentration CI_1 calculé sur cette distribution $\widehat{y^1}$ (selon la méthode utilisant l'estimation (3.2)) donne le degré d'inégalité dû aux inégalités de morbidité. Il se rapproche de l'indice *Need* mentionné en section 3.2.3, car c'est la part de l'indice d'inégalité dû aux différences de besoin de soins selon le revenu.

2) La distribution permettant de mettre en évidence l'effet de l'hétérogénéité des comportements à morbidité donnée (les *pratiques*) relatifs à la participation, selon le groupe q auquel appartient chaque individu, peut s'écrire de la façon suivante (en notant i l'individu et q le quantile de revenu auquel il appartient) :

$$\forall i, q, \widehat{y_{i,q}^2} = G(\overline{X_1}\widehat{b} + M_{1,i}\widehat{d_q}) * F(\overline{X_2}\widehat{\beta} + M_{2,i}\widehat{\delta}) \tag{3.15}$$

Où l'on a remplacé $\widehat{d}$ par $\widehat{d_q}$ dans l'expression (3.14). La différence entre l'indice de concentration CI_2 calculé sur cette distribution $\widehat{y^2}$ et l'indice calculé précédemment CI_1 donne le degré d'inégalité dû aux inégalités de pratiques à morbidité donnée, relatives à la participation.

2bis) La distribution permettant de mettre en évidence l'effet de l'hétérogénéité des comportements à morbidité donnée (les *pratiques*) relatifs à la consommation conditionnelle,

différenciés selon le groupe q auquel appartient chaque individu, peut s'écrire de la façon suivante (en notant i l'individu et q le quantile de revenu auquel il appartient) :

$$\forall i, q, \widehat{y_{i,q}^{2bis}} = G(\overline{X_1}\widehat{b} + M_{1,i}\widehat{d_q}) * F(\overline{X_2}\widehat{\beta} + M_{2,i}\widehat{\delta_q}) \tag{3.16}$$

Où l'on a remplacé $\widehat{\delta}$ par $\widehat{\delta_q}$ dans l'expression (3.15). La différence entre l'indice de concentration CI_{2bis} calculé sur cette distribution $\widehat{y^{2bis}}$ et l'indice calculé précédemment CI_2 donne le degré d'inégalité dû aux inégalités de pratiques à morbidité donnée, relatives à la consommation conditionnelle.

3) La distribution permettant de mettre en évidence l'effet des différences de variables socio-économiques selon les individus (par exemple le revenu) peut s'écrire de la façon suivante (en notant i l'individu et q le quantile de revenu auquel il appartient) :

$$\forall i, q, \widehat{y_{i,q}^{3}} = G(X_{1,i}\widehat{b} + M_{1,i}\widehat{d_q}) * F(X_{2,i}\widehat{\beta} + M_{2,i}\widehat{\delta_q}) \tag{3.17}$$

Où l'on a remplacé $\overline{X_1}$ par $X_{1,i}$ et $\overline{X_2}$ par $X_{2,i}$ dans l'expression (3.16).La différence entre l'indice de concentration CI_3 calculé sur cette distribution $\widehat{y^3}$ et l'indice calculé précédemment CI_2 donne le degré d'inégalité dû aux inégalités socio-économiques.

4) La distribution permettant de mettre en évidence l'effet des différences dans l'impact des variables socio-économiques selon le revenu (ex : différences de l'effet du revenu selon le quantile de revenu auquel on appartient) peut s'écrire de la façon suivante (en notant i l'individu et q le quantile de revenu auquel il appartient) :

$$\forall i, q, \widehat{y_{i,q}^{4}} = G(X_{1,i}\widehat{b_q} + M_{1,i}\widehat{d_q}) * F(X_{2,i}\widehat{\beta_q} + M_{2,i}\widehat{\delta_q}) \tag{3.18}$$

Où l'on a remplacé $\widehat{\beta}$ par $\widehat{\beta_q}$ et $\widehat{b}$ par $\widehat{b_q}$ dans l'expression 3.17. La différence entre l'indice de concentration CI_4 calculé sur cette distribution $\widehat{y^4}$ et l'indice calculé précédemment CI_3 donne le degré d'inégalité dû aux différences d'impact des variables

socio-économiques sur le recours aux soins, selon le revenu.

5) Si l'année en question est instaurée une mesure économique MES dont on cherche à évaluer l'impact sur les inégalités de recours aux soins observées, la distribution de recours aux soins tenant compte de cette mesure se définit par (en notant $\mathcal{M}$ l'ensemble des bénéficiaires de la mesure) :

$$\widehat{y_i^5} = G(X_{1,i}\widehat{b_{MES}} + M_{1,i}\widehat{d_{MES}}) * F(X_{2,i}\widehat{\beta_{MES}} + M_{2,i}\widehat{\delta_{MES}}) \ si \ i \in \mathcal{M} \tag{3.19}$$

$$\widehat{y_i^5} = \widehat{y_i^4} \ sinon \tag{3.20}$$

L'indice de concentration CI_5 est calculé de la même façon que précédemment, et la différence $CI_5 - CI_4$ donne la contribution de la mesure MES à l'inégalité observée.

Remarque : dans le cas où une telle mesure n'est pas implémentée l'année en question, $\forall i, \widehat{y_i^5} = \widehat{y_i^4}$ et $CI_5 = CI_4$.

6) Enfin, la différence entre CI_5 et l'indice de concentration observé C donne le degré d'inégalité dû à des facteurs non-présents dans le modèle (hétérogénéité non-observée).

Pour résumer, C peut se décomposer en la somme des 7 éléments suivants :

- $\Delta_1 = CI_1 - CI_0$ $(= CI_1 - 0 = CI_1)$ est la contribution des inégalités de morbidité
- $\Delta_2 = CI_2 - CI_1$ est la contribution de l'hétérogénéité des pratiques relatives à la participation

- $\Delta_{2bis} = CI_{2bis} - CI_2$ est la contribution de l'hétérogénéité des pratiques relatives à la consommation conditionnelle
- $\Delta_3 = CI_3 - C_{2bis}$ est la contribution des inégalités socio-économiques
- $\Delta_4 = CI_4 - CI_3$ est la contribution de l'hétérogénéité des comportements liés aux variables socio-économiques
- $\Delta_5 = CI_5 - CI_4$ est la contribution de la mesure économique *MES* (le cas échéant)
- $\Delta_6 = C - CI_5$ est la contribution de l'hétérogénéité non-observée

Et on a bien $C = \Delta_1 + \Delta_2 + \Delta_{2bis} + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6$.

3.3.5 Interprétation

La part de C due aux besoins différenciés selon le revenu est égale à Δ_1, que l'on nommera dans la suite *Need*, pour des raisons de comparabilité avec les études existantes (voir équation (3.6)) :

$$Need = \Delta_1. \tag{3.21}$$

La part de C due à d'autres facteurs que les besoins, dénommée *HI*, toujours pour des raisons de comparabilité avec les études existantes, est égale à $\Delta_2 + \Delta_{2bis} + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6$:

$$HI = \Delta_2 + \Delta_{2bis} + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6. \tag{3.22}$$

Soulignons que dans *HI*, nous plaçons également le résidu, c'est-à-dire la part d'inégalité non expliquée par les variables du modèle, ce qui correspond à la pratique habituelle : van Doorslaer et Masseria (2004) incluent également l'hétérogénéité non observée comme facteur explicatif.

La décomposition ici proposée permet de connaître l'effet global de la morbidité sur le degré d'inégalité observé (l'indicateur *Need*), sans avoir de problème de calcul et d'interprétation des C_k présents par exemple dans l'expression (3.4). De plus, alors que dans les décompositions standard l'hétérogénéité des pratiques selon le revenu n'est pas spécifiée, l'hétérogénéité des pratiques est explicitement prise en compte dans notre approche

à travers les coefficients $\widehat{b_q}$ et $\widehat{\beta_q}$. Comme précisé précédemment, les groupes de variables ici choisis suivent la distinction d'usage entre les variables de "besoin" et les variables de "contrôle". Cependant, notre méthode permet de calculer la contribution à l'inégalité de n'importe quel groupe de variable ou même n'importe quelle variable prise isolément, ainsi que de son (ses) coefficient(s).

L'indice *HI* tel qu'il est défini selon l'expression (3.22) peut se décomposer entre la part due à la distribution des variables dans l'échantillon, la part due à l'hétérogénéité des coefficients relatifs à ces variables, et la part due à l'introduction éventuelle d'une mesure économique. Ainsi, la part due aux variables est-elle définie par :

$$Variables = \Delta_3 + \Delta_6 \tag{3.23}$$

Alors que la part de *HI* relative aux comportements différenciés selon le revenu peut-elle s'exprimer de la manière suivante :

$$Comportements = \Delta_2 + \Delta_{2bis} + \Delta_4 \tag{3.24}$$

Le degré d'inégalité dû aux comportements différenciés selon le niveau de revenu se décompose en trois effets :

- Δ_2 : les pratiques différenciées liées à la participation, pouvant être liées à la demande
- Δ_{2bis} : les pratiques différenciées liées à la consommation conditionnelle, pouvant être liées principalement à l'offre et dans une moindre mesure à la demande
- Δ_4 : les paramètres différenciés liés aux variables de contrôle, pouvant être liés à la demande

Cette décomposition permet de distinguer les degrés d'inégalité dûs à l'offre et à la demande de soins, à travers l'analyse de Δ_2 et de Δ_{2bis}, et à travers cette décomposition d'identifier finement les préférences des agents.

3.4 Trois applications à l'analyse de l'inégalité de consommation de soins en France

Les données proviennent de l'Appariement (données IRDES) des années 1998 et 2002 restreint aux individus de 15 ans et plus. Les données comportent respectivement 4336 individus et 4034 individus exploitables, dont respectivement 588 et 691 personnes de 65 ans et plus. Le nombre de bénéficiaires de la CMU-C est de 196 pour l'année 2002, soit 5% de l'échantillon de cette année. Le recours aux soins observé concerne le nombre de consultations de médecins, subdivisé entre séances de généralistes et de spécialistes. Soulignons que l'analyse a été aussi conduite sur les dépenses de pharmacie et d'hôpital : le choix a cependant été fait de ne pas présenter l'intégralité des résultats, afin de focaliser l'étude sur les séances de médecin, dont l'analyse est plus facilement interprétable.
L'intérêt de distinguer généralistes et spécialistes apparaît dans les statistiques descriptives succintes présentées ci-après.

	Tous médecins	Généralistes	Spécialistes
1998	8,3	4,8	3,5
2002	8,6	4,8	3,7

TAB. 3.1: Nombre moyen de consultations par personne, années 1998 et 2002

Nous proposons trois applications. La première décompose les inégalités de consommation de soins pour l'année 1998, et compare les résultats obtenus avec ceux de la littérature. La deuxième propose une évaluation de l'impact de l'introduction de la CMU-C dans l'évolution du degré d'inégalité observée entre les années 1998 et 2002. Enfin, la troisième effectue une décomposition des inégalités de consommation de soins pour l'année 1998, en distinguant les personnes âgées (65 ans et plus) et les personnes plus jeunes (15 à 64 ans), afin de comparer les sources d'inégalité de ces deux sous-groupes de population.

3.4.1 Méthode

On emploie la méthode définie dans la section 3.3 : choix d'un modèle explicatif, définition des quantiles de revenu q, estimation des paramètres, simulation des distributions et enfin calcul des indices de concentration simulés.

Choix du modèle explicatif du recours aux soins

Le type de modèle adopté est un modèle en deux parties indépendantes expliquant d'une part le recours aux soins, d'autre part la consommation conditionnelle. Les variables d'intérêt sont les nombres de séances. Nous avons choisi de modéliser le nombre de séances de médecins plutôt que la dépense de consultations, afin de faciliter l'interprétation des résultats et de permettre la comparabilité avec les études du groupe ECUITY. Les tarifs de consultation variant peu, le nombre de séances donne une idée relativement correcte de la dépense occasionnée. Au total, nous avons donc 3 variables d'intérêt : nombre total de séances de médecins, nombre de séances de généralistes, nombre de séances de spécialistes. Les variables explicatives de la consommation de soins sont sélectionnées selon leur degré de significativité dans une estimation expliquant le nombre total de séances de médecins. On a choisi de ne pas sélectionner les variables différemment pour les séances de spécialistes et de généralistes, afin de pouvoir facilement comparer les analyses pour les trois types de variables. La méthode de sélection des variables est la même que celle décrite dans le chapitre 2. Les indicateurs synthétiques de morbidité ainsi que la présence de couverture complémentaire sont susceptibles d'être endogènes dans les équations estimées : ils sont testés et seul les exogènes sont retenus. Cette sélection des variables a lieu sur tout l'échantillon d'individus âgés de 15 ans et plus. La liste des variables sélectionnées est disponible en annexe C.3.

La participation On modélise la probabilité de recours aux consultations. Le modèle employé est le modèle *Logit*. Les variables explicatives sont sélectionnées suivant leur degré de significativité (test global de nullité simultanée). De plus, les variables pouvant être endogènes sont testées et seules les exogènes sont retenues (suivant en cela la méthodologie à la Rivers et Vuong (1988), voir chapitre 2). En effet, sur une période d'observation de 1 an, certaines variables de morbidité peuvent ne pas être exogènes : le montant de soins peut influencer le degré d'invalidité par exemple. Instrumenter ces variables complexifierait trop la démarche de simulation dans de tels modèles non linéaires : les données étant riches en variables de morbidité, on se contente de supprimer les non-exogènes. Soulignons qu'omettre ces variables non-exogènes ne provoque pas de biais de variable omise : les indi-

cateurs de morbidité sélectionnés dans les estimations sont bien exogènes (leur exogénéité est vérifiée lors des tests).

La modélisation du nombre de séances de praticiens Pour la modélisation du nombre de séances conditionnellement à la participation au système de soins, nous avons choisi ici la loi Binomiale Négative tronquée à 0 (*Zero-Truncated Negative Binomial*). En effet, les données ont une caractéristique qualifiée de *sur-dispersion* : l'espérance de la variable expliquée est supérieure à sa moyenne. Ceci est une caractéristique très souvent retrouvée dans la littérature (voir van Doorslaer, Koolman, et Jones (2004) par exemple). La loi de Poisson caractérisée par une variance égale à la moyenne n'est pas appropriée dans ce cadre (Grogger et Carson 1991). On utilise ici un modèle *GLM* avec pour fonction de lien le *log* et comme loi la loi binomiale négative tronquée à zéro (voir section 1.2.3). Le mode de calcul du nombre de séances prédites est détaillé en annexe C.2. Les variables sont également sélectionnées, suivant la méthodologie développée dans le chapitre 2.

Choix du nombre de quantiles Q Pour l'échantillon couvrant les individus de 15 ans et plus, un choix naturel consiste à choisir les quintiles de revenu, puisque l'échantillon est suffisant. Les individus sont donc classés dans l'ordre croissant de leur revenu (plus précisément, leur revenu par unité de consommation) et pour chaque individu, on définit l'appartenance à tel quintile q, $q \in [1..Q]$, avec $Q = 5$. Lors de l'analyse des personnes de 65 ans et plus, la faiblesse de l'échantillon impose un nombre plus restreint de quantiles de revenu, et dans ce cas $Q = 3$. La comparaison avec le sous-échantillon des individus de 15-64 ans impose d'adopter pour cet échantillon 3 quantiles également.

3.4.2 Application 1 : analyse des inégalités de consommation de soins observées en 1998 en France

Les résultats sont présentés ici sous la forme synthétique de tableaux récapitulant l'inégalité totale, décomposée en plusieurs éléments. Le degré d'inégalité totale, appelé *CI*, est présenté comme la somme de deux éléments, le degré d'inégalité dû au besoin de soins[14] (*Need*) et celui dû à tous les facteurs autres que le besoin (l'indice d'inéquité *HI*). L'indice *HI* est à son tour décomposé en la part due à la distribution des variables selon le revenu, et la part due à l'hétérogénéité des comportements (représentée par l'hétérogénéité des coefficients selon le revenu). Cette part due aux comportements provient de l'hétérogénéité de trois types de coefficients, dont les différents effets sont présentés :

- l'hétérogénéité des comportements de participation liés à la morbidité (ce que nous avons appelé les *pratiques liées à la participation*),
- l'hétérogénéité des comportements de consommation conditionnelle liés à la morbidité (ce que nous avons appelé les *pratiques liées à la consommation conditionnelle*),
- l'hétérogénéité des comportements liés aux autres variables.

Le tableau ci-après fournit la décomposition par facteur de l'inégalité observée en 1998 sur notre échantillon.

	Généralistes	Spécialistes	Tous médecins
CI	**-0,0219**	**0,0884**	**0,0252**
Need	-0,0047	-0,0078	-0,0065
HI	-0,0172	0,0962	0,0317
HI	**-0,0172**	**0,0962**	**0,0317**
Variables	-0,0097	0,0450	0,0135
Comportements	-0,0075	0,0512	0,0182
Comportements	**-0,0075**	**0,0512**	**0,0182**
Participation	0,0004	0,0381	0,0257
Autres	0,0281	0,0481	0,0047
Cons. conditionnelle	-0,0361	-0,0351	-0,0122

TAB. 3.2: Décomposition de l'inégalité observée en 1998 en France

La contribution de chaque facteur a été calculée à partir des indices de concentration des différentes distributions de consommation simulées. En annexe C.4 sont présentés les indices de concentration de chaque distribution simulée et le mode de calcul des valeurs

[14] autrement dit, à la morbidité et à l'âge

présentées dans le tableau 3.2. Une analyse de robustesse à l'ordre d'introduction des effets est également présentée. L'ordre d'introduction des effets dans les simulations décrites en section 3.3.4 ne modifie pas la valeur calculée de ces effets.

Comparabilité avec les résultats de la littérature

Les ordres de grandeur des indices d'inégalité globale *CI* sont proches de ce que l'on trouve dans la littérature. A titre d'exemple, le tableau 3.3 présente la décomposition par facteur de l'inégalité observée sur les séances de généralistes sur données françaises par van Doorslaer et Masseria (2004), à partir de la même enquête que celle que nous avons utilisée, à savoir l'Appariement (année 2000). Les auteurs emploient la méthode décrite en section 3.2. Dans cette étude, l'indice de concentration des séances de généralistes est de $-0,0275$ pour l'année 2000, alors que nous trouvons une valeur de $-0,0219$ sur nos données pour l'année 1998. L'indice total d'inégalité *CI* est décomposé en la somme de la part due aux besoins de soins (*Need*) et la part due à tous les autres facteurs (*HI*), lui même décomposé en différents facteurs. La ligne *Résidu* est ajoutée par nos soins, et représente la différence entre l'indice *CI* observé et la somme de tous les effets identifiés ; pour les auteurs, le résidu fait implicitement partie de *HI*.

	Décomposition
CI	-0,0275
Need	-0,0227
HI	-0,0047
Income	-0,0031
Education	-0,0077
Activity status	0,0087
Region	-0,0073
Insurance	0,0181
CMU	-0,0185
Résidu	-0,0177

TAB. 3.3: Décomposition de l'inégalité de consommation de séances de généralistes, année 2000, par ECUITY

L'indice *Need* trouvé par notre méthode diffère largement de celui trouvé par van Doorslaer et Masseria (2004). Pour les séances de généralistes, ces auteurs aboutissent à $Need = -0,0227$, ce qui constitue quasiment la totalité du CI, alors que nous aboutissons à $Need = -0,0047$ avec notre méthode. Les variables de besoin retenues ne sont en effet pas les mêmes. L'étude de van Doorslaer et Masseria (2004) vise à fournir des comparaisons internationales : les auteurs utilisent des variables d'état de santé auto-attribué

(*self-assessed health*), jointes à des variables indiquant le degré de gêne dans la vie courante. Ces variables sont en effet présentes dans la plupart des enquêtes réalisées dans les pays étudiés par les auteurs. Nous travaillons sur la même enquête utilisée dans le cas de la France par van Doorslaer et Masseria (2004), et la santé auto-attribuée faisait partie des indicateurs de morbidité potentiellement utilisés dans notre analyse. Cependant, la sélection des variables opérée en amont a montré que la santé auto-attribuée était une variable non-significative pour l'équation de participation, et non-exogène pour l'équation de consommation conditionnelle : elle a donc été supprimée des analyses. Nous avons pu reproduire notre analyse en reprenant cette variable en tant qu'indicateur unique du besoin de soins : le *Need* trouvé pour 1998 est alors proche du *Need* trouvé par van Doorslaer et Masseria (2004) pour l'année 2000 : $-0,0254$ contre $-0,0227$. Cela montre que notre méthode aboutit à des résultats cohérents avec la littérature existante, *mais* que la sélection rigoureuse des variables que nous avons menée conduit à une décomposition par facteur largement différente, à cause d'une liste de variables explicatives différentes. Au total, dans nos résultats, la part des inégalités due aux différences de besoins est beaucoup plus faible et la part des inégalités interprétées comme "injustifiable" (pour reprendre la terminologie de Wagstaff et van Doorslaer (2000)) est beaucoup plus importante. En tout état de cause, on observe ici une grande instabilité de l'évaluation du *Need* par rapport au choix des régresseurs.

Nos résultats : l'importance des hétérogénéités de comportement dans l'explication des inégalités

Les indices CI sont de signes différents selon la variable considérée : $CI < 0$ pour les séances de généralistes, et $CI > 0$ pour les séances de spécialistes et le nombre total de séances. Les séances de généralistes sont ainsi concentrées chez les plus pauvres, alors que les séances de spécialistes et le nombre total de séances sont concentrés chez les plus riches. L'indice $Need$ est négatif pour les trois variables : cela montre bien que les besoins sont concentrés dans le bas de la distribution des revenus. Soulignons que les *Need* ne sont pas égaux, alors qu'ils représentent tous les trois une indication de la concentration des besoins selon le revenu. L'indicateur *Need* représente l'indice de concentration de la distribution de consommation de soins qui serait observée si tous les individus étaient soignés de la même

façon, et indépendamment de toutes leurs autres caractéristiques. Cette distribution reflète bien la distribution des besoins selon le revenu, mais dépend de la relation *moyenne* entre besoin et consommation de soins, estimée sur tout l'échantillon. Cette relation diffère selon le type de soin étudié (généralistes, spécialistes, tous médecins) : la distribution simulée $\widehat{y_i^1}$ (équation (3.14) section 3.3.4) diffère donc également, de même que l'indicateur *Need* correspondant. Les résultats sont toutefois cohérents quel que soit le type de soin considéré et indique que les pauvres sont globalement en moins bonne santé que les riches.

L'indice *HI*, qui représente le degré d'inéquité horizontale, est négatif pour les séances de généralistes et positif pour les séances de spécialistes et le nombre total de séances. Cela signifie que pour les séances de généralistes, les plus pauvres reçoivent en proportion *plus* que les riches, compte tenu de leurs besoins. A l'inverse, pour les séances de spécialistes ce sont les riches qui reçoivent en proportion *plus* que les pauvres. Au total, en agrégeant les séances de médecins, l'indice *HI* est positif, indiquant que les plus riches consomment une plus grande proportion de séances.

L'indice *HI* se subdivise entre la partie due à la distribution des variables et la partie due à l'hétérogénéité des comportements. C'est cette partie due aux comportements qui est rendue visible par la méthode que nous développons. On observe que les inégalités dues aux hétérogénéités de comportement sont loin d'être négligeables : chacune équivaut à la moitié de *HI*, et ce pour les trois types de soins. On constate ici l'intérêt d'avoir pu estimer un modèle pour chaque quintile de revenu : la moitié du degré d'inéquité horizontale est due à des comportements différenciés selon le quintile de revenu. Cet effet n'est pas identifiable dans les études employant la méthodologie standard. Dans ces études (cf tableau 3.3), l'indice *HI* se décompose en la somme des effets des variables explicatives *mais* ne tient pas compte de l'effet différencié selon le revenu que ces variables peuvent avoir sur la consommation de soins. D'après l'expression (3.4) (section 3.2), l'indice d'inégalité est en effet décomposé en la somme des indices de concentration des différentes variables explicatives du modèle, pondérés par le coefficient de chaque variable estimé sur tout l'échantillon. Les comportements différenciés selon le revenu ne peuvent donc pas être identifiés. Grâce à notre méthode, nous montrons que ces effets comptent pour moitié

dans l'explication du degré d'inéquité horizontale obtenu.

La part de *HI* due aux comportements se subdivise en l'effet de plusieurs facteurs. L'effet des comportements est globalement pro-pauvre dans le cas des séances de généralistes et pro-riche dans les cas des séances de spécialistes. Les trois composantes de la part de *HI* due aux comportements sont cependant de même signe pour des trois types de soins. L'hétérogénéité des pratiques liées à la participation est clairement pro-riche : à morbidité donnée et à autres caractéristiques données, les plus pauvres tendent à moins recourir aux soins de médecins (généralistes ou spécialistes). En revanche, l'hétérogénéité des comportements liés à la consommation conditionnelle a un caractère pro-pauvre. Une fois la décision prise de consommer, les plus pauvres reçoivent en proportion plus que les plus riches, à morbidité équivalente.

Pour les trois types de consommation, l'effet de l'hétérogénéité des pratiques liées à la participation apparaît pro-riche. Or les comportements de participation à morbidité donnée sont susceptibles de faire partie des comportements de demande. Ce résultat montre la présence d'une hétérogénéité d'accès liée à la position dans l'échelle des revenus. Cette hétérogénéité peut être le signe d'une préférence moindre des personnes les plus pauvres pour la santé. En effet, les coefficients capturant les pratiques à morbidité donnée sont estimés grâce à un modèle où le revenu est présent ainsi que les autres caractéristiques socio-économiques individuelles : les pratiques à morbidité donnée sont ainsi estimées en contrôlant des autres variables non liées au besoin pouvant influencer la consommation de soins.

L'hétérogénéité des pratiques liées à la consommation conditionnelle apparaissent quant à elles pro-pauvre pour tous les types de consommation de soins. Or les comportements déterminant la consommation conditionnelle à morbidité donnée sont susceptibles de faire partie des comportements des offreurs. Le comportement pro-pauvre des offreurs de soins pourrait ainsi compenser les comportements de participation pro-riches des individus : le système de santé jouerait donc bien son rôle en favorisant l'égalité de traitement. Cependant, le moindre recours aux soins de la part des plus pauvres à morbidité don-

née peut impliquer une prise en charge plus tardive, et donc une dégradation accrue de l'état de santé. Ceci peut mener à une consommation plus importante de la part des plus pauvres une fois le recours aux soins décidé, mais également à un résultat moins favorable dû à ce recours tardif. Dormont et Huber (2006) ont montré sur données françaises que lors de l'estimation d'un modèle de consommation de soins en deux parties utilisant la spécification *Sample selection model* (voir section 1.3), le coefficient de corrélation estimé entre les perturbations des deux équations était négatif : à état de santé donné, une plus faible probabilité de participation implique une plus forte consommation de soins en cas de participation. Il est vraisemblable que ce l'on observe ici relève de ce phénomène, qui est largement connu en Santé Publique.

La consommation du nombre total de séances de médecins est globalement inéquitable, et favorise les plus riches ($HI = 0,0317$). La moitié de cette inéquité observée est due à l'hétérogénéité des comportements ($0,0182$). Cette hétérogénéité pro-riche des comportements est principalement due à l'hétérogénéité pro-riche des pratiques liées à la participation, qui sont des comportements que nous pouvons rapprocher des comportements de demande. La question de l'interprétation de ces résultats reste entière à ce stade : ils peuvent être interprétés comme une moindre préférence des pauvres pour la santé, tout comme une indication d'un problème d'accès aux soins lié au revenu. L'interprétation mérite d'être approfondie. Dans l'hypothèse où une préférence moindre pour la santé était mise en évidence chez les plus pauvres, la question du respect de ces préférences peut être soulevée.

3.4.3 Application 2 : impact de la CMU-C dans la variation d'inégalité observée entre 1998 et 2002

La mise en place de la CMU visait à permettre aux plus démunis de se soigner sans être contraints par un budget limité, et donc de réduire les inégalités d'accès aux soins dues au revenu. Certaines études ont montré que les bénéficiaires de la CMU-C consommaient, toutes choses égales par ailleurs, plus que les autres assurés (Raynaud 2003). Outre le fait que les personnes à bas revenu sont statistiquement en moins bonne santé que les autres (par exemple à cause du manque de prévention, de la pénibilité du

travail ...), la question de l'aléa moral des bénéficiaires de la CMU-C a été invoquée pour expliquer cette surconsommation. Une étude de Grignon et Perronnin (2003) a suggéré qu'il pouvait s'agir d'un simple phénomène de calendrier : s'il peut exister un phénomène de rattrapage l'année de l'affiliation à la CMU-C, on ne trouve pas de hausse significative de la dépense moyenne en régime permanent. De plus, les comportements de recours aux soins à morbidité donnée comme les consultations de spécialistes, auxquelles les personnes à faible revenu ont traditionellement peu recours, tendent à converger vers le comportement moyen chez les bénéficiaires de la CMU-C. Les premières conclusions mèneraient donc à penser que la CMU-C remplit son office, c'est-à-dire supprimer les barrières financières à l'accès aux soins pour les personnes les plus pauvres (personnes âgées exceptées). Le nombre moyen de consultations pour les bénéficiaires de la CMU montre une consommation totale plus élevée que la moyenne (ce qui peut être expliqué par leur moins bon état de santé), et largement concentrée vers les séances de généralistes.

	Tous médecins	Généralistes	Spécialistes
Echantillon total	8,6	4,8	3,7
Bénéficiaires de la CMU-C	11,6	7,4	4,2

TAB. 3.4: Nombre moyen de consultations par personne, 2002

La CMU complémentaire a-t-elle contribué à réduire les inégalités de consommation de soins ? Revenons vers le tableau 3.3 : pour van Doorslaer et Masseria (2004), l'effet pro-pauvre de la CMU-C constitue 67% de l'indice de concentration pro-pauvre des séances de généralistes (dans un indice de $-0,0275$, la contribution de la CMU-C est de $-0,0185$). Dans le cas des visites de spécialistes (tableau non reproduit), l'indice de concentration des séances de spécialistes est de $0,0369$ et la contribution de la CMU-C à cet indice est $-0,0166$ (soit 45% en valeur absolue). Couffinhal, Dourgnon, et Tubeuf (2004) confirment ces résultats en montrant que la CMU-C est en effet facteur de réduction des inégalités de consommation de soins pour l'année 2000.

La réduction des inégalités due à l'introduction de la CMU-C n'a apparemment pas été d'égale ampleur pour les séances de généralistes et de spécialistes. Ces dernières

restent concentrées chez les plus riches, malgré l'introduction de la CMU-C.

La mise en œuvre de notre méthode aboutit à la décomposition des inégalités de consommation de soins par facteur pour l'année 2002 donnée par le tableau (3.5). L'année 2002 a été choisie afin que les effets de calendrier mentionnés plus haut soient amortis au moment de nos observations. L'effet de la CMU-C est distingué suivant la méthodologie proposée en section 3.3, et apparaît séparément dans le tableau. L'effet de la CMU-C est inclus dans l'indice HI mais pas dans l'indice captant l'influence des comportements, par souci de comparabilité avec l'application précédente de la section 3.4.2. L'introduction de la CMU-C est en effet une mesure économique introduite à partir de l'année 2000 et l'on souhaite évaluer son effet séparément des autres comportements de recours aux soins. Compte tenu de la méthodologie retenue (consistant à estimer le modèle sur les bénéficiaires de la CMU-C, contrairement aux études standard qui tiennent compte de son effet à travers une indicatrice individuelle), l'effet de la CMU-C peut être assimilée à un changement de comportement. D'après l'équation (3.20) (section 3.3.4), l'effet de la CMU-C est apprécié par un changement dans les coefficients de l'expression employée pour prédire la consommation individuelle.

	Généralistes	Spécialistes	Tous médecins
CI	**-0,0717**	**0,0589**	**-0,0143**
Need	-0,0307	-0,0070	-0,0196
HI	-0,0410	0,0658	0,0053
HI	**-0,0410**	**0,0658**	**0,0053**
Variables	-0,0099	0,0530	0,0211
Comportements	-0,0025	0,0260	0,0067
CMU	-0,0287	-0,0132	-0,0224
Comportements	**-0,0025**	**0,0260**	**0,0067**
Participation	0,0461	0,1120	0,0581
Autres	0,0154	-0,0131	0,0044
Cons. conditionnelle	-0,0640	-0,0730	-0,0557

TAB. 3.5: Décomposition de l'inégalité observée en 2002 en France

La comparaison de ce tableau avec la décomposition obtenue pour l'année 1998 (tableau 3.2) permet de remarquer que les indices de concentration du nombre de séances a diminué pour les trois types de variables : il est même passé de pro-riche à pro-pauvre pour le nombre de séances total. L'indice $Need$ est toujours négatif, indiquant un besoin de soins concentré chez les plus pauvres. Pour les trois variables, l'indice d'inéquité horizontale HI

est de même signe que pour l'année 1998 et est plus faible : on constate donc un mouvement pro-pauvre de cet indice entre les années 1998 et 2002, ce qui signifie une réduction des inégalités favorisant les plus riches. Dans l'indice *HI*, la part de l'effet de la CMU-C est considérable : pour les séances de généralistes, où $HI < 0$, cet effet constitue jusqu'à 70% de *HI*. L'effet de la CMU-C est moins fort pour les séances de spécialistes : alors que $HI > 0$ pour cette variable, l'effet de la CMU-C est négatif et représente 20% en valeur absolue de *HI*. Cet effet réducteur d'inégalités de la CMU-C se retrouve dans van Doorslaer et Masseria (2004) et Couffinhal, Dourgnon, et Tubeuf (2004). Pour van Doorslaer et Masseria (2004), le *HI* est de $-0,0047$ pour les séances de généralistes et de $0,0633$ pour les séances de spécialistes, la contribution de la CMU-C s'élevant respectivement à $-0,0185$ et $-0,0166$: pour les auteurs, l'effet réducteur d'inégalité de la CMU-C est plus important pour les séances de généralistes que de pour les séances de spécialistes, ce que nous retrouvons sur nos données. Notre spécification est cependant plus souple car les auteurs introduisent la CMU-C via une indicatrice, alors que nous estimons un modèle entier sur tous les bénéficiaires de la CMU-C.

On constate que pour l'année 2002, l'effet de la CMU-C est bien pro-pauvre : la part de l'indice dû à la CMU-C est négative. Cet effet de la CMU-C est plus prononcé pour les séances de généralistes que pour les séances de spécialistes. Qu'en est-il cependant de l'impact de la CMU-C sur la réduction de l'inégalité totale observée observée entre 1998 et 2002 ? Le tableau ci-dessous résume l'évolution des indices d'inégalité et d'équité horizontale entre les années 1998 et 2002, ainsi que l'évolution des indices d'inégalité dûs aux besoins (*Need*) et de l'importance de l'effet CMU-C dans ces variations.

	Généralistes	Spécialistes	Tous médecins
Signe de CI 1998 / 2002	- / -	+ / +	+ / -
Δ CI	**-0,0498**	**-0,0295**	**-0,0394**
Δ Need	-0,0260	0,0008	-0,0131
Δ HI	-0,0238	-0,0303	-0,0264
dont effet CMU	-0,0287	-0,0132	-0,0224

TAB. 3.6: Variation de l'inégalité entre 1998 et 2002 : décomposition par facteur

La première ligne du tableau indique le signe de CI observé les années 1998 et 2002. On a $\Delta CI < 0$ et $\Delta HI < 0$: on retrouve le mouvement pro-pauvre mentionné précédemment. On constate que pour les séances de généralistes, de spécialistes, et le nombre total de séances, l'effet de la CMU-C représente respectivement 120%, 43% et 85% de ΔHI : l'effet réducteur des inégalités est très marqué au niveau du nombre total de séances de médecins, mais cet effet est très inégalement réparti selon le type de médecin considéré. Soulignons que l'effet de la CMU-C est important, eu égard au faible effectif des bénéficiaires de la CMU-C présents dans l'échantillon. Une action très ciblée sur la bas de la distribution des revenus aurait donc un effet important au niveau de l'indice de concentration de la consommation de soins. D'autres mesures d'inégalité pourraient donner des résultats différents. En tout état de cause, il conviendrait de mener cette analyse sur un échantillon plus important, afin de vérifier la robustesse des résultats.

3.4.4 Application 3 : comparaison entre les 15-64 ans et les 65 ans et plus, année 1998

Les personnes de 65 ans et plus consomment plus de séances de médecins que les plus jeunes, la différence étant exclusivement due au nombre de séances de généralistes. Peut-on déceler une inéquité de consommation de soins chez les 65 ans et plus ? Et si oui, la décomposition de cette inéquité diffère-t-elle de ce que l'on observe chez les 15-64 ans ? Le chapitre 2 a en effet montré l'importance des pratiques à morbidité donnée dans l'explication des dépenses de santé. L'effet des pratiques hétérogènes selon le revenu se retrouve-t-il chez les personnes âgées ?

		Tous médecins	Généralistes	Spécialistes
65 ans et plus	1998	11,5	7,9	3,6
	2002	11,7	7,7	3,9
15-64 ans	1998	7,8	4,3	3,5
	2002	7,9	4,2	3,7

TAB. 3.7: Nombre moyen de consultations par personne, années 1998 et 2002

La décomposition menée sur le sous-échantillon des 15-64 ans et celui des 65 ans et plus mène aux résultats suivants :

	Généralistes		Spécialistes		Tous médecins	
	15-64	65+	15-64	65+	15-64	65+
CI	**-0,0231**	**-0,0132**	**0,0887**	**0,0879**	**0,0276**	**0,0176**
Need	-0,0055	0,0008	-0,0100	0,0017	-0,0083	0,0016
HI	-0,0175	-0,0140	0,0987	0,0862	0,0359	0,0160
HI	**-0,0175**	**-0,0140**	**0,0987**	**0,0862**	**0,0359**	**0,0160**
Variables	-0,0196	-0,0293	0,0568	0,0729	0,0153	0,0023
Comportement	0,0021	0,0153	0,0419	0,0133	0,0206	0,0137
Comportement	**0,0021**	**0,0153**	**0,0419**	**0,0133**	**0,0206**	**0,0137**
Participation	0,0203	-0,0138	0,0117	0,1001	0,0071	-0,0005
Autres	-0,0178	0,0135	0,0879	-0,1851	0,0259	-0,0235
Cons. conditionnelle	-0,0004	0,0156	-0,0578	0,0984	-0,0124	0,0377

TAB. 3.8: Décomposition des inégalités par facteur : comparaison des 15-64 ans et des 65 ans et plus, année 1998

Pour le nombre de séances total de médecins, l'indice de concentration est pro-riche pour les 15-64 ans et les 65 ans et plus, mais il est légèrement plus élevé pour ces derniers. Alors que pour les séances de spécialistes, l'indice de concentration est équivalent pour ces deux sous-groupes d'individus, il est nettement moins pro-pauvre pour les 65 ans et plus pour les séances de généralistes. L'indice *Need* est beaucoup moins inégalitaire chez les personnes âgées que chez les plus jeunes : l'état de santé apparaît donc relativement indépendant du niveau de revenu pour les 65 ans et plus. Pour les personnes âgées, l'inégalité de consommation de soins ne provient donc pas de différences de morbidité mais plutôt des comportements pouvant être hétérogènes selon le revenu.

L'indice *HI* est négatif pour les séances de généralistes, et son signe est principalement dû à la distribution des variables selon le revenu. L'impact pro-riche des comportements est beaucoup plus fort pour les 65 ans et plus que pour les autres. Cet effet des comportements a des caractéristiques très différentes selon les groupes d'âge considérés : alors que pour les 15-64 ans, les comportements liés à la participation sont pro-pauvre et ceux liés à la consommation conditionnelle pro-riche, on constate l'inverse pour les 65 ans et plus. Le statut d'activité des individus appartenant à ces deux sous-groupes de population peut expliquer ces différences dans les comportements de participation : les 65 ans et plus sont retraités, alors que les 15-64 ans sont actifs dans leur grande majorité. Pour les personnes actives à faible revenu, il peut être plus coûteux de poser un jour de congé que pour une personne à haut revenu, car l'impact de cette perte de salaire peut être beaucoup plus importante en terme d'utilité. Ceci peut expliquer les pratiques pro-riches relatives à

la participation dans le cas des 15-64 ans. Le fait que l'on observe chez les personnes âgées une hétérogénéité pro-pauvre des pratiques relatives à la participation peut être révélateur de l'absence de coût d'opportunité de ces personnes lors de consultations. Ces personnes étant dans leur majorité retraitées, elles ne font pas face à un coût d'opportunité relatif à la nécessité de poser un jour de congé pour une visite chez le médecin généraliste.

Pour les séances de spécialistes, *HI* doit son importance plus à la distribution des variables qu'à l'hétérogénéité des comportements. Cette importance des variables est beaucoup plus marquée chez les 65 ans et plus : l'effet des variables constitue près de 90% de *HI* pour ce sous-groupe de population. Pour ce type de soins, les personnes de 65 ans et plus ont des comportements moins pro-riches que les 15-64 ans (0,0133 contre 0,0419).

Au total, la consommation de séances de médecins apparaît moins inéquitable chez les personnes âgées que chez les personnes plus jeunes. Pour ces deux sous-groupes de population, l'hétérogénéité des comportements est clairement pro-riche. Cependant, la décomposition de ces comportements montrent qu'il sont de nature très différente selon le groupe considéré. Pour le nombre total de séances, l'hétérogénéité des pratiques liées à la participation est pro-riche pour les 15-64 ans et pro-pauvre pour les 65 ans et plus, ce qui semble indiquer que cette hétérogénéité peut être due au statut d'activité des individus en question plus qu'à leur revenu. Le fait que l'hétérogénéité des pratiques liées à la consommation conditionnelle soit pro-pauvre pour les 15-64 ans et pro-riche pour les 65 ans et plus est difficilement interprétable. Tout comme le système de santé peut dans le cas des plus jeunes compenser la participation pro-riche par une consommation conditionnelle pro-pauvre, il est possible d'assister au mécanisme inverse dans le cas des personnes âgées. Le fait que les comportements des offreurs de soins soient discriminatoires envers les personnes âgées les plus pauvres mériterait d'être confirmé par d'autres analyses.

3.5 Conclusion

La méthode d'analyse des inégalités que nous proposons permet une analyse synthétique des différents facteurs à l'origine des inégalités de recours aux soins observées. L'approche par micro-simulations permet d'éviter le calcul et l'interprétation de quantités difficilement interprétables présentes dans les décompositions standard. L'estimation par quantile de revenu permet quant à elle de mettre en évidence l'effet sur les inégalités des hétérogénéités de comportement à morbidité donnée, ce qui est impossible avec les méthodes standard. La présence d'inégalités selon le revenu dans la consommation d'un bien est habituellement testée en comparant les comportements relatifs à différents quintiles de revenu (Wagstaff et van Doorslaer 2000). La mesure de l'effet de cette hétérogénéité des comportements est cependant absente de la décomposition standard des mesures d'inégalité. La décomposition de l'indice de concentration que nous proposons ici permet donc de concilier les analyses contrefactuelles à la Oaxaca (1973) permettant de mettre en évidence des différences de comportement par quantile de revenu, et les analyses visant à décomposer par facteur le degré d'inégalité.

L'étude permet d'analyser les sources d'inégalité dans la consommation de séances de médecins observée. On retrouve sur nos données des résultats connus de la littérature, à savoir qu'à morbidité donnée, les soins de généralistes sont concentrés chez les plus pauvres, et les soins de spécialistes concentrés ches les plus riches.

Les trois applications proposées ont permis de mettre en évidence l'importance de l'hétérogénéité des comportements dans l'inégalité observée. Suivant l'usage de la littérature, l'analyse décompose l'indice d'inégalité de la consommation de soins observée en part "justifiable" (due au besoin de soins, appelée *Need*) et part "injustifiable" (due à tous les facteurs autres que le besoin, appelée *HI*), suivant la terminologie employée par Wagstaff, van Doorslaer, et Watanabe (2003). L'indice d'inéquité horizontale *HI* est décomposé en la part due aux comportements et la part due aux caractéristiques individuelles : l'analyse montre que l'effet de l'hétérogénéité des comportements peut représenter jusqu'à la moitié de l'indice *HI*. La prise en compte de l'hétérogénéité des comportements est donc

fondamentale dans l'explication des sources d'inégalité de consommation de soins. Cette hétérogénéité des comportements peut être liée à l'hétérogénéité des préférences individuelles.

Nous avons montré que la consommation de soins à morbidité donnée est de moins en moins concentrée chez les plus riches entre les années 1998 et 2000. Nous montrons ainsi que l'indice *HI* diminue entre ces deux années, et que cette baisse est en grande partie imputable à l'introduction de la CMU-C. La CMU-C a donc bien joué son rôle, à savoir une réduction des inégalités de consommation de soins.

L'analyse menée séparément sur les 15-64 ans et les 65 ans et plus a montré que l'hétérogénéité des pratiques à morbidité donnée était de nature différente pour ces deux sous-groupes de population. Pour les séances de généralistes, l'hétérogénéité des pratiques liées à la participation est pro-riche pour les 15-64 ans, et pro-pauvre pour les 65 ans et plus, ce qui peut indiquer un coût d'opportunité plus faible chez les 65 ans et plus, lié au statut d'activité (retraité), ou bien éventuellement un effet de génération.

Nous montrons que la décomposition de l'inégalité par facteur est très liée à la définition des variables de besoin. Une réflexion doit donc être menée sur la définition de ces indicateurs de besoin.

Les résultats demandent à être confirmés sur un échantillon plus important, en particulier ceux concernant l'analyse des bénéficiaires de la CMU-C et des personnes âgées, ce qui de plus permettrait l'inférence sur les résultats. L'analyse ouvre cependant de nombreuses pistes relatives à la compréhension et à l'analyse des sources d'inégalité : une réflexion devra être menée au sujet de la place donnée aux préférences dans le domaine de la santé[15], et surtout à la façon de les mesurer. En tout état de cause, la volonté égalitaire du système de santé français n'a pas entièrement porté ses fruits. Des inégalités de consommation de soins subsistent à morbidité donnée. "Les stratégies efficaces de lutte contre les maladies cardio-vasculaires, les cancers et les maladies de société creusent l'écart

[15] On peut trouver une discussion à ce sujet dans Fleurbaey (2006)

entre ceux qui sauront adopter les stratégies nécessaires et les autres" (Cribier 1997), ces "autres" étant souvent les personnes à plus faible revenu.

Conclusion générale

Cette thèse a permis d'étudier les caractéristiques de la consommation de soins au niveau individuel, en employant des méthodes de micro-simulations sur données individuelles françaises (l'Appariement de l'IRDES). L'approche par micro-simulations s'est révélé très fructueuse pour l'analyse des changements de comportement relatifs d'une part à l'augmentation des dépenses de santé, d'autre part à l'analyse des sources d'inégalité de consommation de soins.

Le chapitre 1 fournit un cadre pertinent pour l'analyse économétrique de la demande de soins de santé au niveau individuel. Ce chapitre a permis de sélectionner un modèle explicatif de la consommation de soins proche des données, le *Modified Two-part Model*, employé dans les deux chapitres suivants. Cette spécification en deux équations indépendantes est comparée à d'autres spécifications, et son emploi est justifié.

Les chapitres 2 et 3 proposent tous deux des analyses par micro-simulations fondées sur les techniques d'estimation développées dans le chapitre 1.

Le chapitre 2 propose une méthode originale visant à comparer les effets du vieillissement démographique à d'autres effets influençant la hausse des dépenses de santé : les changements de pratiques du côté de l'offre et de la demande, et les changements de morbidité. Les résultats montrent que le vieillissement démographique ne joue qu'un rôle mineur dans l'explication de la hausse des dépenses de santé observée entre les années 1992 et 2000. Cette hausse des dépenses est principalement due aux changements dans la dépense moyenne par âge. Or ces changements ne sont pas dûs à une détérioration de

l'état de santé des individus, mais aux changements de comportement à morbidité donnée. Les résultats montrent ainsi que les changements de morbidité intervenus entre les années 1992 et 2000 mènent à une réduction de la dépense moyenne par âge, à pratiques constantes. En revanche, à morbidité donnée, les changements de comportement et en particulier les pratiques à morbidité donnée sont à l'origine de la quasi-totalité de la hausse des dépenses de santé observée entre les années 1992 et 2000. Le spectre du vieillissement démographique, souvent évoqué lors de discussions sur les sources du déficit du système d'Assurance-Maladie, est donc brandi à mauvais escient. Les résultats extrêmement probants que l'on trouve sur les dépenses de pharmacie montrent que les changements de comportement à l'origine de la hausse des dépenses sont liés à la consommation conditionnelle et non à la participation : ce ne sont pas les comportements des individus qui sont en cause, mais plus vraisemblablement les comportements des offreurs de soins, largement influencés par le progrès technique dans le cas de la dépense de pharmacie. L'analyse par micro-simulations sur données individuelles en employant des modèles en 2 équations (participation puis consommation conditionnelle) a permis de distinguer les effets d'offre des effets de demande. Le caractère novateur de cette approche est d'avoir pu grâce aux simulations établir une synthèse des effets de changements de comportements d'une part, et des effets de changements de morbidité d'autre part, afin d'évaluer leur impact global sur la hausse de la consommation de soins de santé. Ainsi, nous disposons d'un outil efficace permettant de quantifier les effets respectifs du vieillissement, des changements de morbidité et des changements de comportements. L'effet sur les dépenses du progrès technique est ici explicitement modélisé de manière synthétique. L'application proposée a permis de montrer que ce dernier effet est de loin le plus important dans l'augmentation des dépenses de santé observées.

Alors que le chapitre 2 analyse l'évolution de la dépense moyenne par âge, le chapitre 3 s'intéresse à la répartition de cette dépense selon le niveau de revenu. En effet, le système de santé français est de conception égalitaire. Les effets mis en évidence au chapitre 2 sont des effets moyens, et la question de savoir si la consommation de soins est équitable ou non selon le niveau de revenu permet d'apprécier la performance du système de soins en matière d'égalité d'accès aux soins. L'analyse s'inspire des travaux du

groupe ECUITY, dont l'objet est une comparaison internationale des sources d'inégalité de consommation de soins. Nous proposons une méthode conservant les principes de base des travaux d'ECUITY, en enrichissant l'approche grâce à l'emploi de micro-simulations. L'indice de concentration de la consommation de soins est calculé sur des distributions simulées sous différentes hypothèses. Grâce à cette approche, nous sommes en mesure d'évaluer des effets d'hétérogénéité de comportements qui ne sont pas identifiables dans l'approche standard. Suivant l'usage de la littérature, nous séparons l'indice d'inégalité de la consommation de soins entre la part due à l'hétérogénéité des besoins et la part due à tous les autres facteurs explicatifs de la consommation de soins. Cette dernière part (l'indice d'inéquité horizontale HI) est qualifiée d' "injustifiable" selon Wagstaff, van Doorslaer, et Watanabe (2003). Nous montrons que l'hétérogénéité des comportements peut représenter jusqu'à la moitié de cet indice HI : l'identification de l'hétérogénéité des comportements est donc cruciale dans l'explication des facteurs à l'origine des inégalités de recours aux soins observées. Nous montrons que le choix des indicateurs de besoins est déterminant dans la décomposition des inégalités de recours aux soins. Les applications proposées permettent en particulier d'évaluer l'impact de l'introduction de la CMU-C dans la réduction des inégalités de consommation de soins observée : la CMU-C est à l'origine de la quasi-totalité de cette réduction des inégalités.

Ces résultats ouvrent de nombreuses pistes. Il serait intéressant de répliquer ces analyses sur des échantillons plus importants afin de valider ces résultats, et également de procéder à l'inférence sur les effets calculés. Cette inférence est difficile car les simulations mettent en œuvre des fonctions non linéaires et la façon la plus abordable de procéder à l'inférence est le rééchantillonnage (malheureusement extrêmement coûteux en temps). La faiblesse des échantillons n'a pas permis une analyse aussi fine que nous l'aurions souhaité. On pense en particulier dans le chapitre 2 à l'impossibilité d'estimer des coefficients spécifiques à chaque tranche d'âge, ou dans le chapitre 3 à l'impossibilité d'estimer un modèle expliquant l'adhésion à la CMU-C. Il s'agit de prolongements souhaitables de l'analyse. La grande qualité des données a cependant permis une analyse extrêmement fine des comportements, grâce à la grande richesse de l'information individuelle en particulier concernant la morbidité individuelle. Nous pensons notamment aux tests d'exogénéité des indicateurs

de morbidité que nous avons pu réaliser, et qui sont absents de la quasi-totalité des études portant sur la consommation de soins.

L'approche par micro-simulations a montré son intérêt dans l'analyse de problèmes aussi divers que l'analyse des facteurs à l'origine de l'évolution des dépenses de santé, que l'analyse des sources d'inégalité de recours aux soins. Le rôle central des comportements individuels et de leur hétérogénéité selon le revenu a été mis en évidence. De nombreuses pistes restent à explorer quant à la mesure et à l'interprétation de ces comportements, et notamment les comportements liés aux préférences. Dans le cadre de l'étude des inégalités de consommation de soins (chapitre 3), l'interprétation de l'hétérogénéité des comportements reste à approfondir. De même, l'interprétation des changements de pratiques à morbidité donnée mis en évidence dans le chapitre 2 demande également réflexion.

Appliquer ces méthodes à des données plus précises (notamment les données suisses et canadiennes) permettront d'observer les changements de pratiques liée au progrès technique ; ces données permettent en effet d'observer la diffusion des techniques innovantes, en particulier à l'hôpital. Des données plus nombreuses permettront également de vérifier vers quelles tranches d'âge le progrès technique est orienté. Ces analyses répliquées sur des données internationales permettront également d'observer le rôle des différents systèmes de régulation sur les comportements.

Annexe A

Description des données IRDES

La description des données est inspirée de Aligon, Com-Ruelle, Dourgnon, Dumesnil, Grignon, et Retailleau (2001). Le mode de construction des données est semblable pour toutes les années étudiées. Les changements intervenus dans le mode d'enquête ne concernent que l'extension de la base d'enquête, et sont précisés dans la suite.

La Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (Cnamts), la Caisse Nationale des Professions Indépendantes (Canam) ainsi que la Caisse Nationale de la Mutualité Sociale Agricole (Msa) disposent chacune d'un Echantillon Permanent d'Assurés Sociaux (Epas), qui lui assurent un instrument de suivi de leurs prestations sous la forme de relevés systématiques des dépenses de santé. Pour chaque assuré ainsi que pour ses ayants-droit, ce relevé concerne non seulement les remboursements effectués par la caisse mais également les montants à la charge de l'assuré, et ce pour tous les postes de dépense : consultations de médecins généralistes, spécialistes, kinésithérapie, frais d'hospitalisation etc.

L'Epas est représentatif au 1/1200e des assurés du régime général, mais ses données ne peuvent donner qu'une idée imparfaite de la dépense médicale, et ce pour plusieurs raisons. Si les dépenses sont recensées de manière systématique, donc complète, les renseignements concernant les assurés à l'origine de ces dépenses sont relativement sommaires : on ne dispose que de leur numéro d'identification (crypté) permettant de les rattacher à un ménage, de leur sexe, de leur date de naissance, ainsi que de leur statut eu égard à la

protection sociale (nature du régime, qualité d'assuré principal ou d'ayant-droit). De plus, un assuré en bonne santé et un assuré décédé auront tous deux dans l'EPAS une consommation de soins nulle : les seules données de l'EPAS ne peuvent permettre de distinguer ces deux cas, si la mise à jour de l'échantillon n'est pas faite immédiatement.

C'est donc principalement pour compléter cette base de données que l'IRDES mène l'Enquête Santé et Protection Sociale (ESPS), en interrogeant chaque année un quart des assurés de l'échantillon. Depuis l'année 1998, l'enquête est menée tous les 2 ans sur la moitié de l'échantillon. Jusqu'en 1994, seul l'EPAS relatif à la CNAMTS était interrogé. Les EPAS relatifs à la CANAM et à la MSA sont enquêtés respectivement depuis 1995 et 1996. Les données recueillies concernent le ménage de l'assuré, c'est à dire lui-même et ses cohabitants, qu'ils soient ou non ayants-droit. Pour tous les membres du ménage, on recueille des renseignements socio-démographiques (âge, occupation principale, revenu...), sur la protection sociale (nature du régime de protection sociale, existence d'une couverture complémentaire...), sur la morbidité (maladies existantes, facteurs de risque...) ainsi que la consommation de soins au cours d'un mois (consignée par l'enquêté dans un carnet de soins).

Les données de l'ESPS, comme toutes données d'enquête, peuvent souffrir de biais, en particulier d'attrition (biais lié à la nature même de l'enquête : par exemple, les individus les plus malades sont peu enclins à répondre), ou bien d'erreurs (les individus peuvent ignorer leur statut administratif exact).

C'est la raison pour laquelle l'IRDES a procédé à des appariements entre les données de l'EPAS et celles de l'ESPS : l'EPAS fournissant des données exactes sur la consommation de soins, et l'l'ESPS apportant une information beaucoup plus riche sur les assurés.

Cependant, la complexité de cette tâche provoque une réduction drastique des effectifs disponibles : des 40 000 assurés de l'EPAS, il n'en reste que 5000 dans l'appariement de l'année 1998, qui pourtant regroupe la moitié des membres de l'échantillon permanent

(donc 20 000 personnes) interrogées cette année là. L'effectif des échantillons appariés à partir de 1998 est d'environ 5000 à 6000 personnes, alors que pour les échantillons appariés avant 1998, les effectifs sont de l'ordre de 3000 à 4000 personnes.

Annexe B

Annexes relatives au chapitre 2

B.1 Variables potentiellement utilisées dans les estimations

B.1.1 Description sommaire des variables

- Tranche d'âge
- Sexe
- Statut marital *(matrim)*
- Taille du ménage *(tmenage)*
- Niveau d'études
- Catégorie socio-professionnelle du chef de ménage *(suivant la classification de l'Insee, CSP)*
- Revenu net par unité de consommation, échelle à préciser *(Revenu)*
- Absence de couverture complémentaire *(Ccompl)*
- Les indicateurs de morbidité décrits ci-après

B.1.2 Les groupes d'âge

- 0 : 0-9 ans
- 10 : 10-19 ans
- 20 : 20-29 ans
- 30 : 30-39 ans

- 40 : 40-49 ans
- 50 : 50-59 ans
- 60 : 60-69 ans
- 70 : 70 ans et plus

B.1.3 Caractéristiques socio-démographiques

Statut marital

- 1 : marié ou en couple
- 2 : divorcé ou séparé
- 3 : veuf / veuve
- 4 : célibataire

Niveau d'études

- 1 : non scolarisé, jamais scolarisé, autre
- 2 : maternelle à école primaire
- 3 : collège
- 4 : lycée
- 5 : études supérieures

Catégorie socio-professionnelle (classification de l'INSEE)

- 1 : agriculteurs
- 2 : indépendants
- 3 : professions intellectuelles
- 4 : professions intermédiaires
- 5 : employés
- 6 : ouvriers qualifiés
- 7 : ouvriers non-qualifiés

Revenu net du ménage par unité de consommation

- 1 : appartenant au premier quintile de la distribution (revenus parmi les 20% les plus faibles de la distribution)
- 2 : appartenant aux quintiles 3 et 4 (revenus parmi les 40% du centre de la distribution)
- 3 : appartenant au dernier quintile (revenus parmi les 20% les plus élevés de la distribution)

B.1.4 Les indicateurs de morbidité (IRDES, notice Appariement année 2000)

Leur adéquation au modèle économétrique est testée : leur significativité est testée, ainsi que l'exogénéité des indicateurs de morbidité synthétiques.

Les indicateurs synthétiques de morbidité :

Le degré d'invalidité : (niveaux 5 à 7 regroupés dans l'analyse)

- 0 : pas de gêne
- 1 : gênée, de façon infime
- 2 : peu gêné
- 3 : gêné, mais vit normalement
- 4 : doit diminuer son activité domestique ou professionnelle
- 5 : activité réduite
- 6 : pas d'autonomie domestique
- 7 : alitement permanent

Le risque vital (RV) :

- 0 : aucun indice de risque vital
- 1 : pronostic péjoratif très faible
- 2 : pronostic péjoratif faible
- 3 : risque possible
- 4 : pronostic probablement mauvais
- 5 : pronostic sûrement mauvais

Le niveau d'invalidité ainsi que le niveau de risque vital (probabilité de décès dans les 5 ans) sont élaborés par les médecins codificateur de l'Irdes au vu des questionnaires.

Le nombre de maladies (nbmal) est calculé à partir de la liste de maladies déclarées par l'individu enquêté. Cette liste de maladies est complétée par les médecins codificateurs de l'enquête sur la base d'autres éléments de l'enquête (comme le type de médicament consommé).

La santé auto-ressentie (note) exprimée par le patient est codée par une note allant de 0 (plus faible niveau de santé imaginable) à 10 (meilleur niveau de santé imaginable).

Les maladies chroniques

les maladies chroniques choisies dans l'analyse sont celles qui affectent particulièrement les personnes âgées. Elles ont été sélectionnées suivant la CIM-10 (Classification Internationale des Maladies, 10e édition) avec l'aide de Catherine Sermet, Médecin et Directeur de Recherche à l'Irdes. Comme la CIM est passé de la CIM-9 à la CIM-10 entre 1992 et 2000, nous avons dû pour notre étude utiliser une correspondance entre les deux classifications, correspondance élaborée par l'Irdes. Elle consiste en la correspondance de groupes de maladies (et non de maladies prises individuellement). L'indicatrice de morbidité correspondant à ces groupes code le fait de souffrir d'au moins une des maladies du groupe. Les groupes de maladies choisis sont les suivants :

- diabète
- broncho-pneumopathies chroniques obstructives et emphysème *(BPCO)*
- cardiopathies ischémiques *(cardio)*
- hypertension *(hta)*
- maladies des artères *(arteres)*
- troubles du métabolisme des lipides *(lipides)*

- dépression *(depression)*
- troubles du sommeil *(sommeil)*
- cataracte *(cataracte)*
- arhtrites, arthropathies, dorsopathies *(arthrodorso)*

B.2 La sélection des variables

Pour chaque type de dépense (dépenses de consultations, de pharmacie et d'hôpital) et chaque étape du modèle (participation et consommation conditionnelle), une estimation a été menée, prenant en compte toutes les variables explicatives potentielles, afin de sélectionner celles qui étaient significatives. Les résultats ci-après donnent la liste des variables sélectionnées, ainsi que celles qui ont été éliminées, avec les résultats des tests de nullité jointe des coefficients de ces dernières.

Pour chaque étape de l'analyse (participation et consommation conditionnelle), l'analyse d'exogénéité utilise comme instruments les variables significatives sélectionnées ainsi que certaines variables supplémentaires, décrites ci-après (qui peuvent varier selon l'équation considérée).

Pour les estimations finales et les simulations, seules les variables significatives et exogènes sont conservées. Cette annexe récapitule le processus de choix pour chaque type de dépense considéré : analyse de significativité, puis tests d'exogénéité. On trouvera également pour chaque type de dépense le R^2 de la régression de l'estimation de l'inverse du ratio de Mill ($\hat{\lambda}$) sur les variables de deuxième étape, qui indique leur degré de corrélation, justifiant la validité de l'emploi des GLM.

B.2.1 Dépenses de consultations

Analyse de significativité

Participation (probit)

<u>Sélectionnées</u> : age, revenu, CSP, tmenage, etudes, matrim, hypertension, lipides, nbmal, ccompl

<u>Eliminées</u> : sexef, BPCO, cardio, arteres, depression, sommeil, cataracte, arthrodorso, RV, invalidite, note, diabete

Test de H_0 : les variables éliminées ne sont pas significatives : chi2(12)=15.14, p=0.2339

Dépense conditionnelle

Sélectionnées : age, sexef, CSP, etudes, depression, invalidite, nbmal, note, ccompl

Eliminées : revenu, diabete, BPCO, cardio, hypertension, arteres, lipides, sommeil, cataracte, arthrodorso, RV, tmenage, matrim

Test de H_0 : chi2(16)=20.12, p=0.2148

R^2 du $\hat{\lambda}$ régressé sur les variables de seconde étape : $R^2 = 0.7568$ (le coefficient de corrélation entre les deux est donc $\sqrt{0.7568} = 0.8799$)

Tests d'exogénéité des indicateurs synthétiques de morbidité et de la couverture complémentaire

Participation (probit)

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI

Test de Sargan (validité des instruments) : chi2(3)=3.150, p=0.3691

Instruments faibles : chi2(10)=415.05, p=0.0000

Variables testées : nbmal, ccompl

Variables exogènes : nbmal, ccompl

Test de Rivers & Vuong : chi2(2)=1.26, p=0.5336

Dépense conditionnelle

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI, taille, IMC (normal, surpoids, obésité, la référence étant sous-poids)

Test de Sargan (validité des instruments) : chi2(5)=6.432, p=0.2664

Instruments faibles : chi2(36)=836.59, p=0.0000

Variables testées : invalidite, nbmal, note, ccompl

Variables exogènes : nbmal, note, ccompl

Test d'Hausman : chi2(5)=7.29, p=0.1996

B.2.2 Dépenses de pharmacie

Analyse de significativité

Participation (probit)

Sélectionnées : age, sexef, CSP, tmenage, etudes, matrim, diabete, hypertension, lipides, arthrodorso, nbmal, note, ccompl

Eliminées : revenu, BPCO, cardio, arteres, depression, sommeil, cataracte, RV, invalidite

Test de H_0 : chi2(10)=5.25, p=0.8738

Dépense conditionnelle

Sélectionnées : age, sexef, tmenage, etudes, diabete, hypertension, lipides, depression, arthrodorso, RV, invalidite, nbmal, note

Eliminées : revenu, CSP, ccompl, BPCO, cardio, arteres, sommeil, cataracte, matrim

Test de H_0 : chi2(17)=13.19, p=0.7233

R^2 du $\hat{\lambda}$ régressé sur les variables de seconde étape : $R^2 = 0.7267$ (le coefficient de corrélation entre les deux est donc : $\sqrt{0.7267} = 0.8525$)

Tests d'exogénéité des indicateurs synthétiques de morbidité et de la couverture complémentaire

Participation (probit)

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI

Test de Sargan (validité des instruments) : chi2(2)=1.059, p=0.5889

Instruments faibles : chi2(15)=631.81, p=0.0000

Variables testées : nbmal, note, ccompl

Variables exogènes : nbmal, note, ccompl

Test de Rivers & Vuong : chi2(3)=1.56, p=0.6687

Dépense conditionnelle

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI, taille, IMC (normal, surpoids, obésité, la référence étant sous-poids)

Test de Sargan (validité des instruments) : chi2(5)=7.488, p=0.1868

Instruments faibles : chi2(36)=1040.39, p=0.0000

Variables testées : RV, invalidite, nbmal, note

Variables exogènes : invalidite, nbmal, note

Test d'Hausman : chi2(5)=9.47, p=0.0917

B.2.3 Dépenses d'hôpital

Analyse de significativité

Participation (probit)

Sélectionnées : age, sexef, tmenage, etudes, matrim, diabete, lipides, cataracte, arthrodorso, RV, invalidite, nbmal, note

Eliminées : revenu, CSP, ccompl, BPCO, cardio, hypertension, arteres, depression, sommeil

Test de H_0 : chi2(15)=15.57, p=0.4115

Dépense conditionnelle

Sélectionnées : age, etudes, hypertension, arthrodorso, invalidite, nbmal

Eliminées : sexef, revenu, ccompl, diabete, BPCO, cardio, arteres, lipides, sommeil, cata-

racte, RV, note, tmenage, matrim, CSP, depression

Test de H_0 : chi2(24)=32.57, p=0.1135

R^2 du $\hat{\lambda}$ régressé sur les variables de seconde étape : $R^2 = 0.7579$ (le coefficient de corrélation entre les deux est donc : $\sqrt{0.7579} = 0.8706$)

Tests d'exogénéité des indicateurs synthétiques de morbidité et de la couverture complémentaire

Participation (probit)

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI, taille, IMC (normal, surpoids, obésité, la référence étant sous-poids)

Test de Sargan (validité des instruments) : chi2(5)=5.519, p=0.3558

Instruments faibles : chi2(36)=1426.97, p=0.0000

Variables testées : RV, invalidite, nbmal, note

Variables exogènes : RV, invalidite, nbmal, note

Test de Rivers & Vuong : chi2(4)=3.20, p=0.5249

Dépense conditionnelle

Instruments : Statut du chef de ménage (employé, au chômage, sans activité, la référence étant retraité), fumeur, RMI, taille, IMC (normal, surpoids, obésité, la référence étant sous-poids)

Test de Sargan(validité des instruments) : chi2(7)=7.023, p=0.4265

Instruments faibles : chi2(18)=94.30, p=0.0000

Variables testées : invalidite, nbmal

Variables exogènes : invalidite, nbmal

Test d'Hausman : chi2(2)=0.19, p=0.9115

B.3 Les estimations

	1992	(Obs : 2745)		2000	(Obs : 4307)	
Var	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
age 0-9	5.829	0.170	0.000	5.995	0.311	0.000
age 10-19	5.594	0.189	0.000	5.970	0.341	0.000
age 20-29	5.833	0.186	0.000	5.948	0.327	0.000
age 30-39	5.814	0.188	0.000	5.924	0.328	0.000
age 40-49	5.749	0.183	0.000	5.980	0.324	0.000
age 50-59	5.843	0.181	0.000	5.904	0.326	0.000
age 60-69	5.927	0.183	0.000	5.888	0.331	0.000
age 70+	5.955	0.191	0.000	6.010	0.333	0.000
female	0.178	0.039	0.000	0.237	0.070	0.001
school : never	-	-	-	-	-	-
school : primary	-0.298	0.089	0.001	-0.341	0.139	0.014
school : junior high	-0.171	0.101	0.089	-0.240	0.162	0.138
school : high school	-0.233	0.110	0.034	-0.385	0.174	0.027
school : higher educ	-0.276	0.115	0.017	-0.301	0.175	0.085
household size	-0.032	0.017	0.057	-0.037	0.030	0.215
diabetes	0.481	0.113	0.000	0.206	0.221	0.351
hypertension	0.401	0.071	0.000	0.528	0.123	0.000
lipid	0.259	0.089	0.003	0.409	0.149	0.006
depression	0.373	0.083	0.000	0.025	0.157	0.875
arthro-dorso	-0.100	0.060	0.095	-0.448	0.101	0.000
nb diseases	0.062	0.012	0.000	0.047	0.022	0.032
Self-Assessed Health	-0.129	0.014	0.000	-0.121	0.027	0.000
disab 0	-	-	-	-	-	-
disab 1	-0.014	0.075	0.856	0.024	0.126	0.851
disab 2	0.203	0.084	0.016	0.415	0.121	0.001
disab 3	0.418	0.092	0.000	1.135	0.143	0.000
disab 4	0.768	0.111	0.000	1.655	0.182	0.000
disab 5	0.886	0.161	0.000	1.390	0.272	0.000
No compl cov	-0.250	0.065	0.000	-0.035	0.116	0.760

TAB. B.1: Equation de consommation conditionnelle GLM, dépenses pharmaceutiques : 1992 and 2000

Test de significativité globale du modèle : $p < 0.0000$ pour 1992 et 2000.

	1992	(Obs : 3441)		2000	(Obs : 5003)	
Var	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
age 0-9	2.304	0.273	0.000	2.459	0.260	0.000
age 10-19	1.829	0.287	0.000	1.934	0.268	0.000
age 20-29	1.509	0.263	0.000	1.595	0.249	0.000
age 30-39	1.459	0.256	0.000	1.647	0.242	0.000
age 40-49	1.368	0.253	0.000	1.388	0.239	0.000
age 50-59	1.367	0.254	0.000	1.197	0.239	0.000
age 60-69	1.494	0.264	0.000	1.452	0.247	0.000
age 70+	1.620	0.297	0.000	1.788	0.279	0.000
diabetes	0.560	0.281	0.046	0.205	0.245	0.403
hypertension	0.442	0.143	0.002	0.422	0.128	0.001
lipid	0.596	0.219	0.006	0.344	0.167	0.039
arthro-dorso	0.264	0.098	0.007	0.133	0.075	0.075
nb diseases	0.039	0.016	0.018	0.073	0.016	0.000
Self-Assessed Health	-0.039	0.020	0.055	-0.043	0.020	0.032
female	0.099	0.053	0.061	0.160	0.048	0.001
soc : farmers	-0.070	0.241	0.772	-0.068	0.194	0.725
soc : indep.	0.108	0.123	0.379	-0.364	0.098	0.000
soc : intell.	0.003	0.100	0.977	0.001	0.075	0.992
soc : interm.	0.290	0.072	0.000	0.071	0.071	0.318
soc : clerks	0.136	0.082	0.096	0.012	0.079	0.881
soc : qual.	-	-	-	-	-	-
soc : non-qual.	0.174	0.096	0.070	0.018	0.090	0.842
school : never	-	-	-	-	-	-
school : primary	-0.305	0.131	0.020	-0.353	0.119	0.003
school : junior high	-0.136	0.143	0.343	-0.169	0.128	0.185
school : high school	-0.252	0.155	0.104	-0.114	0.137	0.406
school : higher educ	-0.448	0.162	0.006	-0.438	0.138	0.001
matrim : married	-	-	-	-	-	-
matrim : divorced	0.105	0.185	0.570	0.095	0.137	0.488
matrim : widow	-0.007	0.200	0.973	-0.090	0.184	0.625
matrim : single	-0.417	0.093	0.000	-0.409	0.083	0.000
Household size	-0.100	0.021	0.000	-0.034	0.020	0.083
No compl cov	-0.403	0.076	0.000	-0.359	0.071	0.000

TAB. B.2: Equation de participation, dépenses pharmaceutiques : 1992 and 2000

Test de significativité globale du modèle : $p < 0.0000$ pour 1992 et 2000.

B.4 Les simulations sur population observée

Les effectifs de population utilisés pour les simulations macro pour les années 1992 et 2000 proviennent de la base Ecosanté de l'IRDES (IRDES 2002).

	1992	2000
Population des <4 ans - Effectifs	3741630	3628054
Population des <1 an - Effectifs	738226	752349
Population des 1-4 ans - Effectifs	3003404	2875706
Population des 5-9 ans - Effectifs	3804951	3620994
Population des 10-14 ans - Effectifs	3880572	3839762
Population des 15-19 ans - Effectifs	3974842	3919982
Population des 20-24 ans - Effectifs	4285315	3729478
Population des 25-29 ans - Effectifs	4318386	4129864
Population des 30-34 ans - Effectifs	4306594	4235153
Population des 35-39 ans - Effectifs	4265736	4344303
Population des 40-44 ans - Effectifs	4356832	4243832
Population des 45-49 ans - Effectifs	3398624	4193473
Population des 50-54 ans - Effectifs	2766273	4074642
Population des 55-59 ans - Effectifs	2920710	2782969
Population des 60-64 ans - Effectifs	2933701	2684772
Population des 65-69 ans - Effectifs	2677852	2704137
Population des 70-74 ans - Effectifs	1964271	2456364
Population des 75-79 ans - Effectifs	1400567	2099952
Population des 80-84 ans - Effectifs	1270523	960018
Population des 85-89 ans - Effectifs	683983	813379
Population des 90-94 ans - Effectifs	239166	341810
Population des 95 ans et + - Effectifs	49325	90050

TAB. B.3: Effectifs de la population française par tranche d'âge, 1992 et 2000

Annexe C

Annexes relatives au chapitre 3

C.1 Les données

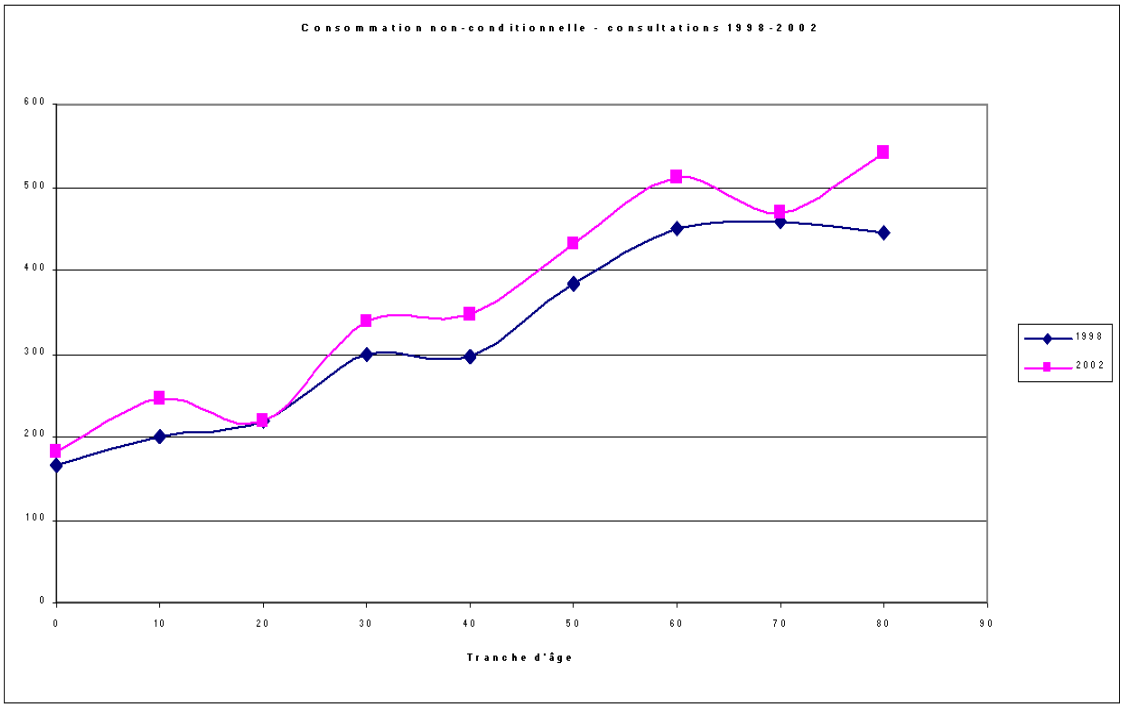


Fig. C.1: Consommation observée en 1998 et 2002 par groupe d'âge, consultations

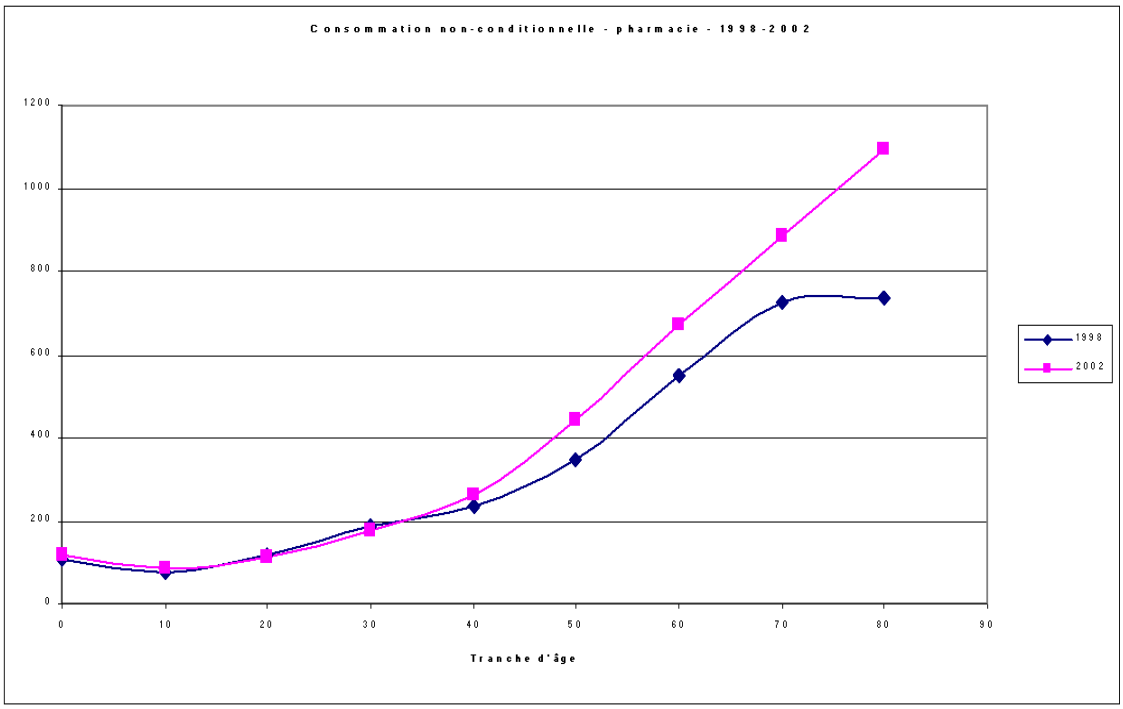


FIG. C.2: Consommation observée en 1998 et 2002 par groupe d'âge, pharmacie

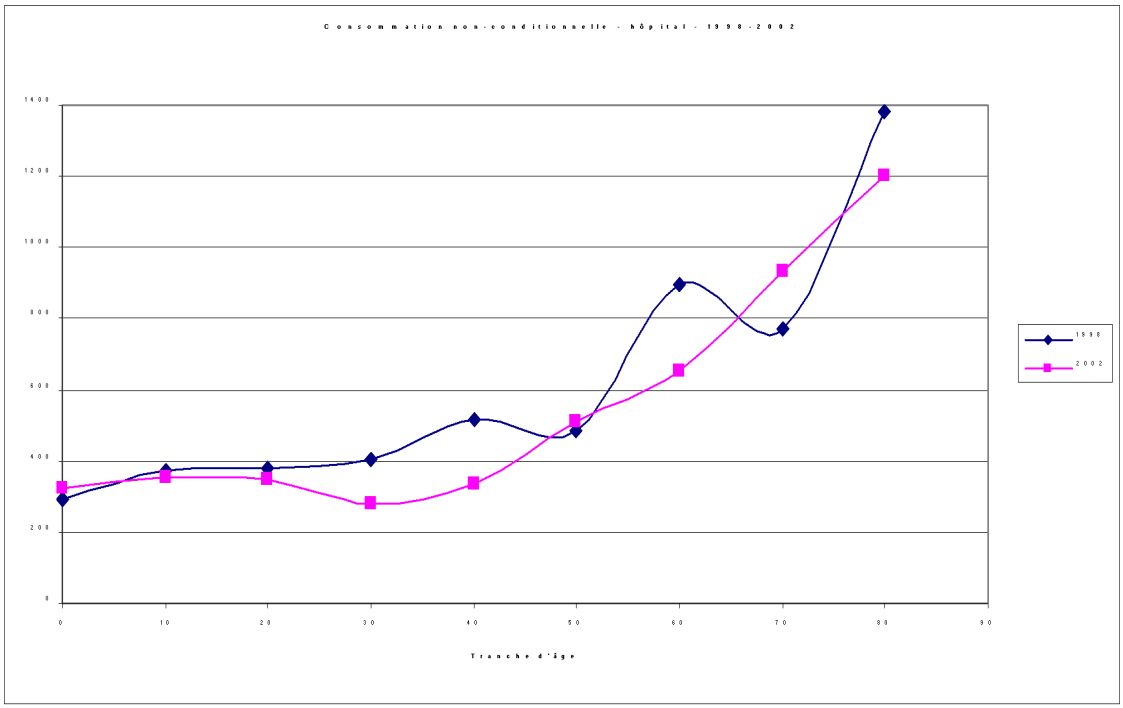


FIG. C.3: Consommation observée en 1998 et 2002 par groupe d'âge, hôpital

C.2 Calcul du nombre de séances conditionnel

Une loi binomiale négative de moyenne μ a pour densité la fonction suivante (Grogger et Carson 1991) :

$$f(y) = P[Y = y] = \frac{\Gamma(y+\theta)}{\Gamma(y+1)\Gamma(\theta)}(\frac{\mu}{\theta})^y(1+\frac{\mu}{\theta})^{-\theta(y+1)} \tag{C.1}$$

Avec :

$$E[Y] = \mu \tag{C.2}$$

$$V[Y] = \mu(1+\frac{\mu}{\theta}) \tag{C.3}$$

Où θ est le paramètre de sur-dispersion, et où la fonction Γ se définit par $\forall\alpha$, $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-t}dt$.

D'après C.1, on a, puisque $\Gamma(1) = 1$:

$$P(Y > 0) = 1 - P(Y = 0) \tag{C.4}$$

$$= \frac{\Gamma(\theta)}{\Gamma(1)\Gamma(\theta)}(\frac{\mu}{\theta})^0(1+\frac{\mu}{\theta})^{-\theta} \tag{C.5}$$

$$= 1 - (\frac{\theta}{\theta+\mu})^\theta \tag{C.6}$$

Par C.6, la loi binomiale négative tronquée à zéro a pour densité :

$$h(y) = P[Y = y|Y > 0] = \frac{P[Y = y]}{P[Y > 0]}\frac{\Gamma(y+\theta)}{\Gamma(y+1)\Gamma(\theta)}(\frac{\mu}{\theta})^y(1+\frac{\mu}{\theta})^{-\theta(y+1)}\frac{1}{1-(\frac{\theta}{\theta+\mu})^\theta} \tag{C.7}$$

L'approche par GLM modélise μ :

$$\mu = \exp(Xb) \tag{C.8}$$

Les paramètres b ainsi que le paramètre de surdispersion θ sont estimés par maximum de vraisemblance, où la vraisemblance du modèle s'écrit à partir de l'expression

C.7.

On a :

$$E[Y|Y > 0] = \frac{\mu}{P(Y > 0)} \tag{C.9}$$

La moyenne d'une loi Binomiale de moyenne μ et de paramètre de sur-dispersion θ tronquée à zéro s'écrit donc :

$$E(Y|Y > 0) = \frac{\mu}{1 - \frac{\theta}{(\theta + \mu)^{\theta}}}. \tag{C.10}$$

L'espérance conditionnelle de y s'écrit donc :

$$E(y_i|P_i = 1) = \frac{\exp(Xb)}{1 - \frac{\theta}{(\theta + \exp(Xb))^{\theta}}} \tag{C.11}$$

Et cette expression s'estime par l'expression suivante, qui est utilisée lors des simulations :

$$\hat{E}(y_i|P_i = 1) = \frac{\exp(Xb)}{1 - \frac{\hat{\theta}}{(\hat{\theta} + \exp(Xb))^{\hat{\theta}}}} \tag{C.12}$$

Avec θ l'inverse du paramètre de sur-dispersion, et $\hat{\theta}$ un estimateur de θ trouvé lors de l'estimation. Lors des simulations de distributions, le $\hat{\theta}$ employé lors du calcul est celui de la régression sur tout l'échantillon et non celui calculé lors de la régression sur le quantile de revenu en question, ceci pour des raisons de comparabilité.

C.3 Les variables retenues

C.3.1 Variables potentiellement utilisées dans les estimations

Le détail de ces variables est présent dans les annexes dédiées au chapitre 2, car il s'agit de la même enquête.

– Age

- Csp du chef de ménage
- Niveau d'études
- Revenu
- Couverture complémentaire
- Nombre de personnes du ménage
- Fait de vivre en couple
- Sexe féminin
- Fait d'être au chômage
- Fait d'être veuf
- Risque vital
- Invalidité
- Note auto-attribuée sur l'état de santé, sur 10
- Diabète
- Broncho Pneumopathies Chroniques Obstructives et Emphysème (BPCO)
- Cardiopathies ischémiques
- Hypertension artérielle
- Maladie des artères
- Cancer
- Trouble du métabolisme des lipides
- Troubles mentaux
- Dépression
- Troubles du sommeil
- Cataracte
- Arthropathies, arthrites, dorsopathies
- Insuffisance rénale
- Nombre de maladies

L'âge est pris en compte à travers des indicatrices codant les dizaines d'années (10-19 ans, 20-29 ans ...), jusqu'à la tranche 70 ans et plus. La Csp du chef de ménage est codée selon la classification de l'Insee, en regroupant les Csp 1 et 2, les Csp 4 et 5, et les Csp 6 et 7. Le degré d'invalidité est codé en 3 indicatrices, regroupant respectivement les degrés 0 et 1, 2 et 3, et 4 et plus. Le risque vital est également codé en 3 indicatrices,

regroupant respectivement les degrés 0 et 1, 2 et 3, et 4 et plus. Le nombre de maladies est codé en 3 indicatrices, regroupant respectivement un nombre de maladies compris entre 0 et 3, 4 et 7, et un nombre de maladies égal à 8 et plus. La note auto attribuée sur l'état de santé est codée grâce à 3 indicatrices : 0-5 (note faible), 6-7 (note moyenne), 8 et plus (note élevée). Les variables codant pour les pathologies chroniques codent pour la présence d'au moins une maladie parmi la famille de maladie qu'elles regroupent.

C.3.2 Variables sélectionnées dans les estimations du chapitre 3

Equation de participation :

- Age
- Sexe féminin
- Catégorie socio-professionnelle du chef de ménage
- Revenu (Euros)
- Fait de vivre en couple
- Couverture complémentaire
- Hypertention artérielle
- Troubles mentaux
- Degré d'invalidité
- Nombre de maladies

Equation de consommation conditionnelle :

- Age
- Sexe féminin
- Revenu (Euros)
- Nombre de personnes du ménage
- Fait de vivre en couple
- Fait d'être au chômage
- Fait d'être veuf
- Diabète
- Broncho Pneumopathies Chroniques Obstructives et Emphysème (BPCO)
- Cardiopathies ischémiques

- Cancer
- Troubles mentaux
- Dépression
- Cataracte
- Degré d'invalidité
- Nombre de maladies

C.4 Les distributions simulées

C.4.1 Identification des sources d'inégalités de consommation de soins : l'approche par microsimulations

Cette annexe présente les résultats relatifs aux deux premières applications, c'est-à-dire l'analyse des sources d'inégalité en France en 1998 et 2002, pour les individus de 15 ans et plus. Nous nous restreignons ici aux séances de généralistes. La présentation de l'intégralité des tableaux de résultats étant trop lourde, tous les autres tableaux sont disponibles sur demande. Le principe des calculs est le même pour toutes les applications proposées.

Indice de concentration des distributions simulées

Le mode de simulation des distributions est décrit en section 3.3.4. Les indices de concentration de chacune de ces distributions sont calculés sur ces distributions successives. Pour l'année 1998, les indices de concentration sont présentés pour les séances de généralistes dans le tableau C.1.

	1998		2002	
	0		0	
CI0 (égalitaire)				
IC (95%)	-	-	-	-
CI1 (morbidité)	-0,0056		-0,0243	0
IC (95%)	-0,0139	0,0027	-0,0338	-0,0148
				0
CI2 (age)	-0,0047		-0,0307	0
IC (95%)	-0,0139	0,0045	-0,0415	-0,0199
				0
CI3 (prat part)	-0,0043		0,0154	0
IC (95%)	-0,0147	0,0062	0,0032	0,0277
				0
CI4 (prat cond)	-0,0403		-0,0486	0
IC (95%)	-0,0506	-0,03	-0,0619	-0,0353
				0
CI5 (socio)	-0,0509		-0,0688	0
IC (95%)	-0,0617	-0,04	-0,0833	-0,0542
				0
CI6 (coef socio)	-0,0227		-0,0533	0
IC (95%)	-0,0326	-0,0128	-0,065	-0,0417
				0
CI7 (cmu)	-0,0227		-0,082	0
IC (95%)	-0,0326	-0,0128	-0,0943	-0,0697
				0
CI observé	-0,0219		-0,0717	0
IC (95%)	-0,0403	-0,0036	-0,0924	-0,051

Tab. C.1: Indices de concentration des distributions simulées - généralistes, 15+

Les indices de concentration sont calculés grâce à la régression (3.2) présentée dans le chapitre 3. Les intervalles de confiance proviennent de cette régression : ils sont ainsi calculés comme si les distributions étaient certaines ; or elles sont elles-même simulées : les intervalles de confiance présentés ne tiennent pas compte de la variabilité due aux simulations.

Calcul des effets des différentes sources d'inégalité

	Δ 98	Δ 02
Morbidité	-0,0056	-0,0243
Age	0,0009	-0,0064
Prat part	0,0004	0,0461
Prat cond	-0,0361	-0,064
Socio	-0,0105	-0,0202
Coef socio	0,0281	0,0154
Cmu	0	-0,0287
Heterog	0,0008	0,0103

Les différents effets Δ sont calculés comme stipulé en section 3.3.4.

C.4.2 Analyse de la robustesse à l'ordre d'introduction des effets

Le tableau suivant présente, pour les séances de spécialistes, les indices de concentration des distributions simulées pour l'année 1998.

	1998	
CI0 (égalitaire)	0	
IC (95%)	-	-
CI1 (morbidité)	-0,0034	
IC (95%)	-0,0092	0,0024
CI2 (age)	-0,0078	
IC (95%)	-0,0126	-0,0029
CI3 (prat part)	0,0303	
IC (95%)	0,0239	0,0367
CI4 (prat cond)	-0,0047	
IC (95%)	-0,0148	0,0054
CI5 (socio)	0,049	
IC (95%)	0,035	0,063
CI6 (coef socio)	0,0971	
IC (95%)	0,0882	0,1061
CI7 (cmu)	0,0971	
IC (95%)	0,0882	0,1061
CI observé	0,0884	
IC (95%)	0,0661	0,1107

Le tableau suivant présente quant à lui les indices de concentration des distributions simulées, où l'ordre d'introduction des effets a été modifié : l'hétérogénéité des pratiques a été introduite en premier, et ce n'est qu'ensuite qu'est introduite l'hétérogénéité d'état de santé.

	1998	
CI0 (égalitaire)	0	
IC (95%)	-	-
CI1 (prat cond)	-0,0436	
IC (95%)	-0,0459	-0,0414
CI2 (prat part)	-0,0217	
IC (95%)	-0,0235	-0,0199
CI3 (age)	-0,0255	
IC (95%)	-0,0293	-0,0217
CI4 (morbi)	-0,0277	
IC (95%)	-0,0342	-0,0211
CI5 (socio)	0,0357	
IC (95%)	0,0228	0,0486
CI6 (coef socio)	0,0999	
IC (95%)	0,0912	0,1086
CI observé	0,0884	
IC (95%)	0,0658	0,111

L'impact de chaque étape à l'indice d'inégalité total est calculé comme stipulé dans la section 3.3.4. La comparaison de ces effets est présentée dans le tableau suivant :

	Δ 98
Morbidité	-0,0034
Age	-0,0044
Prat part	0,0381
Prat cond	-0,0351
Socio	0,0537
Coef socio	0,0481
Heterog	-0,0087

Tab. C.2: Apport de chaque étape à l'inégalité - spécialistes, 15+

	Δ 98
Prat cond	-0,0436
Prat part	0,0219
Age	-0,0038
Morbi	-0,0022
Socio	0,0633
Coef socio	0,0642
Heterog	-0,0115

Tab. C.3: Apport de chaque étape à l'inégalité - spécialistes 15+

Après permutation de l'ordre d'introduction des effets, leur évaluation est de même ordre de grandeur.

Bibliographie

Ai, C., et E. Norton (2000) : "Standard errors for the retransformation problem with heteroscedasticity," *Journal of health economics*, 19, 697–718.

Aligon, A., L. Com-Ruelle, P. Dourgnon, S. Dumesnil, M. Grignon, et A. Retailleau (2001) : "La consommation médicale en 1997 selon les caractéristiques individuelles," Rapport Irdes, Série Résultats.

Bac, C., et G. Cornilleau (2002) : "Comparaison internationale des dépenses de santé : une analyse des évolutions dans sept pays depuis 1970," *Etudes et Résultats, DREES*, 175.

Basu, A., W. Manning, et J. Mullahy (2004) : "Comparing alternative models : log vs. Cox proportional hazard ?," *Health economics*, 13, 749–765.

Baubeau, D., F. Bousquet, et M. Joubert (2001) : "Le traitement chirurgical de la cataracte en France," *Etudes et Résultats, DREES*, 101.

Bien, F. (2001) : "Essais en Economie de la Santé et Assurance," Thèse de Doctorat, Université Paris 10 Nanterre.

Blough, D., C. Madden, et M. Hornbrook (1999) : "Modeling risk using generalized linear models," *Journal of health economics*, 18, 153–171.

Bourguignon, F., F. H. G. Ferreira, et P. Leite (2002) : "Beyond Oaxaca-Blinder : Accounting for Differences in Household Income Distribution Across Countries," DELTA Working Paper.

Box, G., et D. Cox (1964) : "An analysis of transformations," *Journal of the Royal Statistical Society, series B*, 26, 211–246.

Brouard, N. (2005) : "Eléments de démographie," Séminaire LEGOS, Université Paris Dauphine.

Cameron, A., et P. Trivedi (2006) : *Microeconometrics - methods and applications*. Cambridge University Press.

CNRS (2006) : "Le Trésor de la langue Française informatisé," `http://atilf.atilf.fr/tlf.htm`.

Couffinhal, A., P. Dourgnon, et S. Tubeuf (2004) : *Changes in inequalities in the utilization of ambulatory care over the last decade in France*. Ihea Congress Poster session.

Cribier, F. (1997) : "L'allongement de la vie et la mortalité différentielle des travailleurs après la retraite : l'étude d'une cohorte de parisiens," *Sociologie Santé*, 16, 21–38.

Cutler, D. (2005) : "The Potential For Cost Savings In Medicare's Future," *Health Affairs*, p. hlthaff.w5.r77, `http://content.healthaffairs.org/cgi/content/abstract/hlthaff.w5.r77v1`.

Cutler, D., et E. Meara (2003) : *Perspectives on the economics of aging*chap. Changes in the age distribution of mortality over the 20th century. University of Chicago Press.

Cutler, D., et L. Sheiner (2001) : *Demographic Change and Fiscal Policy*chap. Demographics and the Demand for Medical Spending : Standard and Non-Standard Effects, pp. 253–291. Cambridge University Press.

Delattre, E. (2000) : "Comportements d'offre de soins des médecins français," Thèse de Doctorat, Université Paris 10 Nanterre.

Dobson, A. (1993) : *An introduction to generalized linear models*. Chapman and Hall, London.

Dormont, B. (1999) : *Introduction à l'économétrie*. Montchrestien.

Dormont, B., M. Grignon, et H. Huber (2006) : "Health expenditure growth : reassessing the threat of aging," *Health Economics*, 15, 947–963.

Dormont, B., et H. Huber (2006) : "Causes of health expenditures growth : the predominance of changes in medical practices over population ageing," A paraître dans Annales d'Economie et Statistiques `http://www.dauphine.fr/eurisco/eur\_cr/cah0603.pdf`.

Dow, W. H., et E. C. Norton (2002) : "The red herring that eats cake : Heckit versus two-part model redux," Triangle health economics working paper series - www.unc.edu/the.

Duan, N. (1983) : "Smearing estimate : a nonparametric retransformation method," *Journal of the American Statistical Association*, 78, 605–610.

Felder, S., M. Meier, et H. Schmitt (2000) : "Health care expenditures in the last months of life," *Journal of Health Economics*, 19, 679–695.

Fleurbaey, M. (2006) : "Health, Equity and Social Welfare," *Annales d'Economie et de Statistiques*, à paraître ; document de travail CERSES : `http://cerses.shs.univ-paris5.fr/publifleurbaey/heasw.pdf`.

Fries, J. (1980) : "Ageing, natural death, and the compression of morbidity," *New England Journal of Medicine*, 303, 130–135.

Gerdtham, U., et B. Jönsson (2000) : *Handbook of Health Economics*vol. 1, chap. International comparisons of health expenditures, pp. 11–53. Elsevier.

Gerdtham, U. G., B. Jonsson, M. MacFarlan, et H. Oxley (1998) : *Health, the medical profession and regulation*chap. The determinants of health expenditure in the OECD countries : a pooled data analysis, pp. 113–134. Kluwer Academic Publisher.

Gerdtham, U. G., J. Sogaard, F. Andersson, et B. Jonsson (1992) : "An econometric analysis of health care expenditure : a cross-section study of the OECD countries," *Journal of Health Economics*, 11, 63–84.

Getzen, T. E. (1992) : "Population aging and the growth of health care expenditures," *Journal of Gerontology*, 47, S98–S104.

Gourieroux, C. (1999) : *Mesures d'inégalités de pauvreté et de concentration*. Polycopié ENSAE 2e année.

Gourieroux, C., et A. Monfort (1993) : *Handbook of Statistics*vol. 11, chap. Pseudo Maximum Likelihood Methods, pp. 335–362. Elsevier.

——— (1995) : *Statistique des Modèles Econométriques*. Economica.

Gourieroux, C., A. Monfort, et A. Trognon (1984) : "Pseudo-maximum likelihood methods : applications to Poisson models," *Econometrica*, 52, 701–720.

Greene, W. (2002) : *Econometric Analysis, 5e édition*. Prentice Hall.

Grignon, M. (2002) : "Impact macro-économique du vieillissement de la population sur les dépenses d'assurance-maladie en France," *Santé Société et Solidarité*, 2, 135–154.

Grignon, M., et M. Perronnin (2003) : "Impact de la CMU-C sur les consommations de soins," *Questions d'économie de la santé, IRDES*, 74.

Grogger, J., et R. Carson (1991) : "Models for truncated counts," *Journal of Applied Econometrics*, 6(3), 225–238.

Grossman, M. (1999) : "The human capital model of the demand for health," *NBER Working Paper*, W7078.

Heckman, J. J. (1979) : "Sample selection bias as a specification error," *Econometrica*, 47, 153–161.

Hitiris, T., et J. Posnett (1992) : "The determinants and effects of health expenditure in developed countries," *Journal of Health Economics*, 11, 173–181.

Hogan, C., J. Lunney, J. Gabel, et J. Lynn (2001) : "Medicare beneficiaries' costs of care in the last year of life," *Health Affairs*, 20, 188–195.

Hummel, C. (2003) : *L'avenir - Attentes, projets, (dés)illusions, ouvertures*chap. La vieillesse, un âge obsolète ? L'improbable projet des alchimistes du XXIe siècle. Réalités Sociales, Lausanne.

Jusot, F. (2003) : "Revenu et mortalité : analyse économique des inégalités sociales de santé en France," Thèse de Doctorat, EHESS.

Kramer, M. (1980) : "The rising pandemic of mental disorders and associated chronic diseases and disabilities," *Acta Psychiatrica Scandinavica*, 62, 282–287.

le Cacheux, J., et V. Touzé (2003) : "Vieillissement et richesse des nations," *Revue de l'OFCE*, 86, 231–275.

Leu, R. E. (1986) : *Public and private health services : complementarities and conflicts*chap. The public-private mix and international health care costs, pp. 41–63. Blackwell Scientific Publications.

Leung, S. F., et S. Yu (1996) : "On the choice between sample selection and two-part models," *Journal of Econometrics*, 72, 197–229.

L'Horty, Y., A. Quinet, et F. Rupprecht (1997) : "Expliquer la croissance des dépenses de santé : le rôle du niveau de vie et du progrès technique," *Economie et Prévision*, 129-130.

Lubitz, J. (2005) : "Health, Technology, And Medical Care Spending," *Health Affairs*, p. hlthaff.w5.r81, `http://content.healthaffairs.org/cgi/content/abstract/hlthaff.w5.r81v1`.

Lubitz, J., et R. Prihoba (1984) : "The Use of Medicare Services in the Last Two Years of Life," *Health Care Financing Review*, 5, 117–131.

Lubitz, J., et G. Riley (1993) : "Trends in medicare payments in the last year of life," *The New England Journal of Medicine*, 328(15), 1092–1096.

Manning, W., N. Duan, et W. Rogers (1987) : "Monte Carlo evidence on the choice between sample selection and two-part models," *Journal of Econometrics*, 35, 59–82.

Manning, W., J. Newhouse, N. Duan, E. Keeler, et A. Leibowitz (1987) : "Health insurance and the demand for medical care : evidence from a randomized experiment," *The American Economic Review*, 77(3), 251–277.

Manning, W. G. (1998) : "The logged dependent variable, heteroscedasticity, and the retransformation problem," *Journal of Health Economics*, 17, 283–295.

Manning, W. G., et J. Mullahy (2001) : "Estimating log models : to transform or not to transform ?," *Journal of Health Economics*, 20, 461–494.

Manton, K. (1982) : "Changing concepts of morbidity and mortality in the elderly population," *The Milbank Memorial Foundation Quaterly / Health and Society*, 0, 183–244.

Marazzi, A., et C. Ruffieux (1999) : "The truncated mean of an asymmetric distribution," *Computational Statistics and Data Analysis*, 32, 79–100.

Marazzi, A., et V. Yohai (2004) : "Adaptively truncated maximum likelihood regression with asymmetric errors," *Journal of statistical planning and inference*, 122, 271–291.

McCullagh, P., et J. Nelder (1989) : *Generalized linear models, 2e édition*. Chapman and Hall, London.

McGrail, K., B. Green, M. Barer, R. Evans, C. Hertzmann, et C. Normand (2000) : "Age, costs of acute and long-term care and proximity to death : evidence for 1987-88 and 1994-95 in British Columbia," *Age and Ageing*, 29, 249–253.

Mizrahi, A., et P. Sabourdin (1997) : "Enquête décennale sur la santé et les soins médicaux," France.

Mullahy, J. (1998) : "Much ado about two : reconsidering retransformation and the two-part model in health econometrics," *Journal of Health Economics*, 17, 247–281.

Nelder, J., et R. Wedderburn (1972) : "Generalized linear models," *Journal of the Royal Statistical Society*, Series A 135, 370–384.

Newey, W., et K. West (1994) : "Automatic lag selection in covariance matrix estimation," *Review of economic studies*, 61(4), 631–653.

Oaxaca, R. (1973) : "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693–709.

Oberlin, P., M.-C. Mouquet, et T. Folliguet (2004) : "Le traitement invasif des maladies coronariennes," *Etudes et Résultats, DREES*, 289.

OCDE (2006) : "Projecting OECD health and long-term care expenditures : what are the main drivers ?," OECD economics department working papers.

O'Connell, J. (1996) : "The relationship between health expenditures and the age structure of the population in OECD countries," *Health Economics*, 5, 573–578.

O'Neill, C., L. Groom, A. Avery, D. Boot, et K. Thornhill (2000) : "Age and proximity to death as predictors of GP care costs : results from a study of nursing home patients," *Health Economics*, 9, 733–738.

Park, R. (1966) : "Estimation with heteroscedastic error terms," *Econometrica*, 34(4).

Perronnin, M., L. Rochaix, et S. Tubeuf (2006) : "Construction d'un indicateur continu d'état de santé agrégeant risque vital et incapacité," *Questions d'Economie de la Santé, Irdes*, 107.

Piketty, T. (2004) : *L'économie des inégalités*. Repères - La Découverte.

Raynaud, D. (2003) : "L'impact de la CMU sur la consommation individuelle de soins," *Etudes et Résultats, DREES*, 229.

Riley, G., J. Lubitz, R. Prihoba, et E. Rabey (1987) : "The Use and costs of Medicare services by cause of death," *Inquiry*, 24, 233–244.

Rivers, D., et Q. Vuong (1988) : "Limited information estimation and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

Robine, J.-M., et P. Mormiche (1994) : "Estimation de la valeur de l'espérance de vie sans incapacité en France en 1991," *Solidarité Santé*, 1, 17–36.

Salas, C., et J. Raftery (2001) : "Econometric issues in testing the age neutrality of health care expenditures," *Health Economics*, 10, 669–671.

Sandier, S., D. Polton, V. Paris, et S. Thomson (2002) : *Health care systems in eight countries : trends and challenges*chap. France, pp. 31–45. The European observatory on health care systems.

Sauvy, A. (1953) : "Un exemple pour la Wallonie, La renaissance démographique française," *Revue du Conseil économique wallon*, 2.

Sermet, C. (1998) : "Vieillissement et grandes pathologies : quelles évolutions récentes ?," *Les cahiers de l'INED : la santé aux grands âges*, 147, 115–128.

Seshamani, M., et A. M. Gray (2004a) : "Aging and health-care expenditure : the red herring argument revisited," *Health Economics*, 13, 303–314.

——— (2004b) : "A longitudinal study of the effects of age and time to death on hospital costs," *Journal of Health Economics*, 23, 217–235.

Staiger, D., et J. Stock (1997) : "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

Stearns, S. C., et E. C. Norton (2004) : "Time to include time to death ? The future of health care expenditure predictions," *Health Economics*, 13, 315–327.

Stooker, T., J. van Acht, E. van Barneveld, R. van Vliet, B. van Hout, D. Hessing, et J. Bussbach (2001) : "Costs in the last year of life in the Netherlands," *Inquiry*, 38, 73–80.

IRDES (2002) : "Eco-Santé," Paris, France.

OCDE (1987) : "Financing and delivering health care : a comparative analysis of OECD countries," Rapport OCDE.

van Doorslaer, E., X. Koolman, et A. Jones (2004) : "Explaining income-related inequalities in doctor utilisation in Europe," *Health Economics*, 13, 629–647.

van Doorslaer, E., et C. Masseria (2004) : "Income-related inequality in the use of medical care in 21 OECD countries," Equipe Française : A. Couffinhal , S. Tubeuf , P. Dourgnon , Irdes.

van Doorslaer, E., A. Wagstaff, H. V. D. Burg, T. Christiansen, D. D. Graeve, I. Duchesne, U.-G. Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Hakkinen, J. John, J. Klavus, R. Leu, B. Nolan, O. O'Donnell, C. Propper, F. Puffer, M. Schellhorn, G. Sundberg, et O. Winkelhake (2000) : "Equity in the delivery of health care in Europe and the US," *Journal of Health Economics*, 19, 553–583.

Wagstaff, A., et E. van Doorslaer (2000) : *Handbook of Health Economics*vol. 1, chap. Equity in health care finance and delivery, pp. 1803–1862. Elsevier.

Wagstaff, A., E. van Doorslaer, et N. Watanabe (2003) : "On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam," *Journal of Econometrics*, 112, 207–223.

Wedderburn, R. (1974) : "Quasi likelihood functions, generalized linear models, and the Gauss-Newton Method," *Biometrika*, 61, 439–447.

WorldBank (2005a) : "The concentration index," `http://siteresources.worldbank.org`.

——— (2005b) : "Measuring inequity in health service delivery," `http://siteresources.worldbank.org`.

Yang, Z., E. Norton, et S. Stearns (2003) : "The Real Reasons Older People Spend More," *Journal of Gerontology : Social Sciences*, 58B(1), S1–S10.

Zweifel, P., S. Felder, et M. Meiers (1999) : "Ageing of the population and health care expenditure : a red herring ?," *Health Economics*, 8, 485–496.

——— (2001) : "Reply to : Econometric issues in testing the age neutrality of health care expenditures," *Health Economics*, 10, 673–674.

Résumé

Vieillissement, dépenses de santé et inégalités de recours aux soins : essais de micro-économétrie appliquée

par

Hélène HUBER

Université Paris X Nanterre,

Brigitte DORMONT, Directrice de Recherche

L'objet de cette thèse est l'analyse des comportements de consommation de soins et de leur impact sur l'évolution et la répartition de la dépense de santé dans la population. Nous menons une analyse sur données individuelles et mettons en œuvre des méthodes originales de micro-simulations afin de mettre en évidence les effets de l'hétérogénéité des comportements au niveau individuel.

Un premier chapitre expose les méthodes économétriques employées lors de l'estimation d'équations explicatives de la consommation de soins sur données individuelles.

Un deuxième chapitre développe une méthode originale visant à comparer les effets du vieillissement démographique à d'autres effets influençant la hausse des dépenses de santé entre les années 1992 et 2000. Nous montrons que la hausse des dépenses est principalement due aux changements de comportement face à la maladie, ces changements de comportements pouvant être en grande partie liés au progrès technique. Le vieillissement démographique n'a qu'un rôle mineur dans l'explication de la hausse des dépenses de santé. De plus, nous montrons que l'état de santé moyen s'améliore.

Un troisième chapitre propose une méthode novatrice de décomposition des inégalités de consommation de soins par facteur. A morbidité donnée, la consommation de soins est concentrée chez les individus les plus riches de la distribution ; or nous montrons que cette inéquité est pour moitié due à l'hétérogénéité des comportements due à la position des individus dans l'échelle des revenus, cette hétérogénéité n'étant pas identifiable dans les décompositions standard. Trois applications sont proposées : la décomposition des inégalités par facteur pour la France en 1998, l'évaluation de l'impact de la CMU-C sur l'évolution des inégalités de consommation de soins, enfin la comparaison des sources d'inégalité entre les personnes âgées (65 ans et plus) et les autres.

Mots-clefs : Dépense de santé - Econométrie - Micro-simulations - Vieillissement - Inégalités

Classification JEL : C15 - C34 - D63 - I11 - I18 - J11