

# Composante redistributive du système d'assurance maladie : une approche par estimation et microsimulation des dépenses de santé sur cycle de vie.

Valerie Albouy\*, Laurent Davezies† and Thierry Debrand‡

March 3, 2009

Version provisoire, ne pas citer

## Abstract

Pour mieux comprendre le rôle de notre système d'assurance maladie, il est important d'avoir une meilleure appréhension des mécanismes redistributifs en jeu. En effet, observer les transferts financiers liés à l'assurance maladie à un moment donné ne permet pas de distinguer la " redistribution intra-personnelle " qui joue un rôle d'assurance entre des périodes de bonne santé et de maladie et la redistribution entre groupe d'individus ( " redistribution interpersonnelle " ) qui vient corriger certaines inégalités de risques entre groupes sociaux. Ces mécanismes redistributionnels restent mal connus car ils nécessitent une prise en compte de l'ensemble des consommations et des cotisations.

Le modèle DESTINIE offre d'intéressantes possibilités dans cette perspective. Le " cœur " du modèle construit des trajectoires socio-démographiques, professionnelles et les revenus à la retraite des individus sont simulées sur tout leur cycle de vie. Les prélèvements à l'assurance maladie sont pour l'essentiel conditionnés par la situation des individus sur le marché du travail et leur revenu, et il est possible d'imputer des dépenses de santé dans le modèle en fonction de caractéristiques observables ce qui permet alors d'établir des bilans sur cycle de vie au niveau individuel. L'idée est donc de définir s'il y a des " gagnants " et des " perdants " dans le système (par exemple : Sont-ce les riches ou les pauvres, les hommes ou les femmes etc. ?).

Pour mettre au point un modèle d'imputation des dépenses de santé sur le cycle de vie, il est nécessaire de recourir à des estimations sur des panels (observations des transitions, prise en compte de l'hétérogénéité individuelle inobservée et de la dépendance d'état...). Nous utilisons l'enquête sur la santé et la protection sociale (ESPS) appariée avec l'échantillon permanent des assurés sociaux (EPAS) sur la période 2000-2005 pour estimer les dépenses hospitalières et ambulatoires selon le genre. Après avoir rapidement présenté les vifs débats des économistes de la santé cherchant à estimer correctement des équations de dépenses, nous proposerons une méthode d'estimation d'un modèle dynamique des dépenses de santé adapté à notre objectif de simulation.

---

\*Institut National de la Statistique et des Etudes Economiques (INSEE)

†Institut National de la Statistique et des Etudes Economiques (INSEE)

‡Institut de Recherche et de Documentation en Economie de la Santé (IRDES); 10, rue vauvenargues  
75018 PARIS email : debrand@irdes.fr

# 1 Motivation

## 1.1 Objectifs de l'étude

Depuis les origines, le système public d'assurance maladie français repose sur un principe simple: chacun cotise selon son revenu et est soigné selon son état de santé ce qui induit deux fonctions distinctes :

- Une fonction assurantielle : les individus cotisent à chaque période (avec des cotisations plus élevées pendant les périodes d'activité) et en contre-partie leurs dépenses de santé sont (plus ou moins partiellement) remboursées. Le système mutualise donc des risques et permet à chaque individu d'opérer des transferts entre des périodes de bonne santé et des périodes de mauvaise santé. En l'absence d'une telle mutualisation, les individus averse au risque seraient contraints de constituer une épargne de précaution plus importante que leur cotisation.
- Une fonction redistributive : les cotisations et prélèvements des individus dépendent (pour l'essentiel) de leur situation sur le marché du travail. Dans le système d'assurance public, la tarification au risque n'est pas celle qui est appliquée. Cet écart entre tarification réelle et tarification au risque se justifie par une aversion pour les inégalités face à la richesse et face au risque médical. Il existe donc des transferts redistributifs des "gros cotisants à faibles risques" vers les "petits cotisants à gros risques".

On a cependant peu de certitude concernant l'intensité et l'orientation de cette redistribution. Les besoins de documentation sur le sujet s'expriment régulièrement, comme **par exemple...**

L'analyse des transferts de l'assurance santé a déjà été étudié en coupe, par exemple par Marical (Marical, 2007). Une analyse en coupe met en évidence que les actifs cotisent beaucoup et consomment peu à la différence des retraités âgés qui consomment plus mais cotisent moins. A âge donné, les personnes ayant les plus hauts revenus cotisent plus, on pourrait donc s'attendre à une redistribution verticale des riches vers les pauvres. Mais les riches ont une espérance de vie plus importante que les pauvres et aux grands âges les dépenses de santé sont élevées alors que les cotisations sont faibles, la redistribution sur le cycle de vie pourrait donc être bien différente que celle observée en coupe. Dans ce cadre, il apparaît donc que les analyses en coupe sont partiellement limitées. Les différences d'espérance de vie selon milieux sociaux ainsi que la forte influence de l'âge sur le montant des cotisations et des dépenses rendent nécessaire une analyse sur le cycle de vie.

Pour mener une telle analyse sur le cycle de vie, il faut disposer des deux éléments : les prélèvements au titre du financement du système public d'assurance maladie et des remboursements. A partir de ces éléments, il faut procéder à une analyse redistributive. De manière très générale la redistribution peut-être appréhendée en comparant le total des remboursements actualisés sur le cycle de vie avec le total des cotisations actualisées sur le cycle de vie, et l'on procède à une analyse "coût-bénéfice" du système pour chaque groupe social. L'incertitude liée au risque médical peut s'appréhender comme une dispersion des consommations médicales (au sein des groupes sociaux). La qualité assurancielle du système peut s'appréhender comme une dispersion des dépenses remboursées au sein des groupes sociaux.

Les remboursements ou la consommation médicale elle-même constituent deux variables d'intérêt pertinentes dans cette perspective. Dans le cas de l'étude des remboursements on s'intéresse plutôt à la couverture du risque médical, dans le cas de l'étude des consommations on s'intéresse au risque lui-même. Les deux aspects de la question sont naturellement liés. Nous avons donc choisi de travailler dans un premier temps sur la consommation de soin plus que sur les remboursements pour différentes raisons :

- Il nous semblait important de faire d'un premier temps une analyse descriptive du risque médical sur le cycle de vie avant d'analyser sa couverture par l'assurance maladie.
- La principale difficulté dans une analyse sur cycle de vie est de bien prendre en compte l'hétérogénéité entre individus qui persiste au cours de la vie, tout en respectant des statistiques descriptives en coupe. Les études économétriques disponibles se sont principalement focalisées l'estimation des consommations (généralement en coupe), plus que sur l'estimation des remboursements. Il était donc plus facile de comparer les estimations que nous menions avec les résultats de la littérature.
- Même si le but final est de produire des bilans redistributifs, donc de raisonner sur des montants remboursés et non consommés, il est facile de simuler des montants remboursés une fois la simulation des consommations effectuée dans Destinie. Pour cela soit nous simulerons les montants remboursés soit sur la base de la consommation en estimant un taux de couverture, soit nous reprendrons la même modélisation et la même méthode de simulation mais en changeant la variable dépendante (on modélise et on simule les montants remboursés et non consommés).

La source sur laquelle nous travaillons nous permet d'observer les dépenses de santé des individus affiliés à la CNAMTS, au régime social des indépendants (RSI) et à la mutualité sociale agricole (MSA). Les ayants droits sont également suivis, ainsi que les individus couverts par un régime spécial mais en ménage avec un "individu CNAMTS-RSI-MSA". Seuls les ménages dont tout les membres sont couverts par les régimes spéciaux échappent à l'enquête. Le champs couvert par cette source est donc proche de 97 % de la population française.

Même si il est prévu à terme d'implémenter des dépenses hospitalières dans Destinie pour pouvoir mener une analyse complète, nous nous contentons ici de présenter une analyse concernant les dépenses ambulatoires (c'est à dire consultations médicales chez les généralistes et spécialistes et dépenses en pharmacie).

## 1.2 Choix adoptés

Il n'existe pas (à notre connaissance) de source de données permettant d'observer simultanément les consommations, les remboursements et les cotisations et cela sur le cycle de vie.

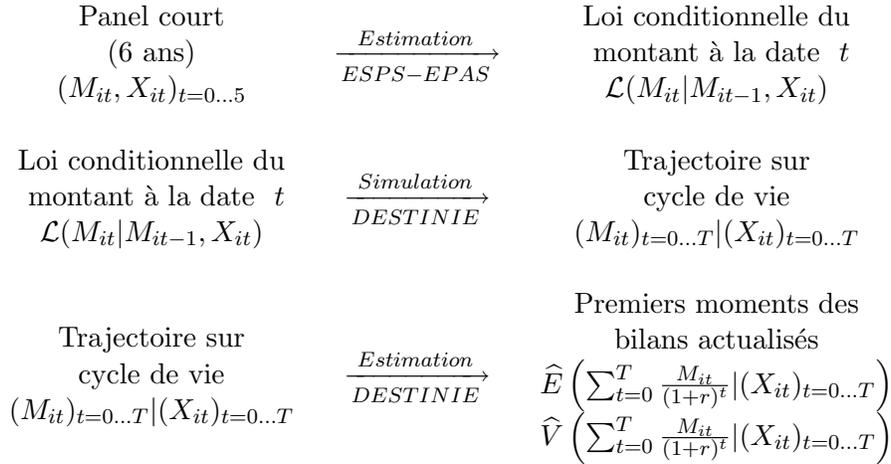
Par contre le modèle Destinie nous fournit un échantillon représentatif de ménages de la population française en 2003. Les individus composants ces ménages et leur descendance sont suivis tout au long de leur vie, avec des trajectoires démographiques et professionnelles reconstruites sur la base des comportements observés actuellement (naissances, études,

mises en couple, décès) et sur la base de prévisions (mortalité, taux de chômage, etc.). A partir de cette modélisation des trajectoires démographiques et professionnelles, il est assez facile d'imputer des prélèvements en utilisant des règles comptables de la législation.

Pour comparer les consommations (ou les remboursements de l'assurance maladie) sur le cycle de vie au niveau individuel, il suffit donc d'être capable de simuler les montants associés dans Destinie. Cela revient peu ou prou à être capable de simuler les dépenses de santé. Pour étudier la redistribution entre grands groupes sociaux, il est important d'avoir une bonne estimation de la dépense moyenne (conditionnellement à des covariables) ; pour étudier le "risque santé", il est important d'avoir une bonne estimation de la diversité des montants de consommation (dispersion, asymétrie etc.); pour étudier la qualité assurantielle du système, il est important d'avoir une bonne estimation de la diversité des remboursements (ici encore dispersion et asymétrie sur les montants remboursés et non consommés). Cela a plusieurs implications :

- Il faut disposer de données permettant d'observer les consommations et remboursements ainsi que le maximum de covariables présentes dans Destinie.
- Si on note  $M_{it}$  les consommations (resp. les remboursements) de l'individu  $i$  à la date  $t$  et  $X_{it}$  des covariables comme l'âge, le sexe, la position sociale etc., la mesure du risque moyen par groupe social (ou les bilans redistributifs entre groupes sociaux si on s'intéresse ou remboursements) feront intervenir la quantité  $E\left(\sum_t \frac{M_{it}}{(1+r)^t} | (X_{it})_t\right)$ , à défaut de disposer d'un panel long contenant l'ensemble des consommations sur le cycle de vie, on peut estimer cette quantité grâce à des estimations en coupe du montant moyen dépensé (ou remboursé) conditionnellement aux covariables :  $E(M_{it} | (X_{it})_t)$ . Pour le dire autrement, pour chaque individu  $i$  à la date  $t$ , il suffit d'imputer une variable aléatoire d'espérance  $E(M_{it} | (X_{it})_t)$  pour obtenir une estimation sans biais des dépenses ou des remboursements sur cycle de vie.
- Les quantités  $E\left(\sum_t \frac{M_{it}}{(1+r)^t} | (X_{it})_t\right)$  ne sont cependant pas les seuls paramètres intéressants. Par exemple, la variance conditionnelle  $V\left(\sum_t \frac{M_{it}}{(1+r)^t} | (X_{it})_t\right)$  est une mesure de la dispersion des dépenses sur le cycle de vie. Si  $M_{it}$  représente le montant consommé, c'est une mesure naturelle et intuitive du caractère difficilement prévisible des dépenses de santé sur cycle de vie, donc de l'incertitude liée au "risque maladie". Si  $M_{it}$  représente le montant remboursé, c'est une mesure simple de la fonction assurantielle de l'assurance maladie : plus les dépenses sont dispersées sur le cycle de vie (conditionnellement aux covariables  $X$ ), plus la fonction assurantielle de l'assurance maladie est importante (transfert entre des individus ayant les mêmes caractéristiques observées  $X$  mais des dépenses de santé différentes).

Notre démarche se décompose en trois grandes étapes :



## 2 Données

### 2.1 Les spécificités ESPS – EPAS

L'étude est réalisée à partir de l'exploitation des données des Enquêtes sur la Santé et la Protection Sociale (ESPS) de l'IRDES appariée avec les données administratives provenant des Echantillons Permanents d'Assurés Sociaux (EPAS) de la Caisse Nationale d'Assurance Maladie (CNAM), du Régime Social des Indépendants (RSI ex-CANAM) et de la Mutualité Sociale Agricole (MSA). Nous utilisons comme base de référence ESPS 2000, d'où nous extrayons les données socio-économiques dont nous avons besoin pour nos estimations des dépenses. L'ESPS est une enquête où les individus répondent à un questionnaire précis décrivant à la fois leurs caractéristiques socio-économiques mais aussi leur santé (état, consommation, ...). Nous avons constitué un panel à partir de ESPS 2000 et des consommations de santé (montant et recours) provenant de l'EPAS pour 2000-2001-2002-2003-2004-2005. Il s'agit d'un panel d'environ 7 112 individus suivis sur 6 années.

Un certain nombre de variables socio-économiques sont invariantes dans le temps : le sexe, le niveau d'éducation. D'autres varient de façon mécanique telles que l'âge. D'autres variables telles que le statut d'occupation et l'état matrimonial ne sont connues qu'en 2000. Nous faisons l'hypothèse que celles-ci restent constantes au cours de la période, nous les prendrons donc comme des invariants issus de ESPS 2000. Les variables de consommation de soins quant à elles varient. Elles diffèrent année après année puisqu'elles proviennent des fichiers EPAS 2000-2001-2002-2003-2004-2005.

Les phénomènes d'attrition peuvent être de différentes natures, les individus peuvent changer de régime de sécurité sociale (hors des trois suscités, par exemple devenir fonctionnaire), ne plus consommer de soin ou être décédés. Le motif d'un refus de réponse par vague est exclu dans notre cas, car les données panélisées sont issues des fichiers administratifs EPAS. Les décès ne sont pas parfaitement identifiables, or nous avons besoin de les détecter pour ne pas considérer de "fausses dépenses nulles" dans nos données. Un certain nombre de décès sont directement identifiés par la CNAM, d'autre part nous avons considéré que les personnes ayant un risque vital élevé en 2000 ainsi que de fortes consommations qui disparaissaient des fichiers EPAS étaient décédés. Pour les dernières

Table 1: Construction

	2000	2001-2003	2004	2005
	ESPS-EPAS	EPAS	ESPS-EPAS	EPAS
Caractéristiques Socio-économiques	Age Sexe CSP Niveau d'étude Statut d'occup. Carac. du ménage		Motif de l'attrition pour la moitié des individus interrogés dans ESPS 2000	
Dépenses de santé	Ambulatoire Hôpital	Ambulatoire Hôpital	Ambulatoire Hôpital	Ambulatoire Hôpital

Table 2: Population

	Effectifs	ambulatoire		Nb de décès
		$D = 1$	$\bar{m}$	
		en €		
2000	7 112	0,92	932	60
2001	7 052	0,93	1 006	43
2002	7 009	0,94	1 105	60
2003	6 949	0,95	1 168	72
2004	6 877	0,92	1 214	45
2005	6 832	0,9	1 222	65
Ensemble	41 831	0,93	1107	345

années, nous avons donc mobilisé les fichiers EPAS 2006 et 2007. Enfin, la moitié de l'échantillon est sélectionné dans l'enquête ESPS 2004 et nous avons pour ces individus un motif de non réponse à l'enquête de 2004 (le décès est fait partie des motifs possibles). Avec ces règles simples nous arrivons à un nombre de décès proche de ce qui est attendu (345 décès sur la période d'observation soit 0.8 % par an).

## 2.2 Statistiques descriptives

Notre objectif dans cette présentation est de construire des équations mobilisables pour estimer des consommations annuelles de soins dans le modèle de micro-simulation DESTINIE de l'INSEE. Dès lors, nous avons dû restreindre le champ de nos variables explicatives. Nous ne retiendrons que celles qui sont actuellement prises en considération dans ce dernier. Nous nous concentrons dans cette présentation sur les dépenses ambulatoires (les dépenses hospitalières sont donc laissées de côté). Nous disposons de deux indicateurs : le premier indique si les gens ont consommé ou pas au cours de l'année ; le deuxième

correspond au montant de la consommation. 93 % des individus de notre panel ont une consommation ambulatoire pour un montant moyen de 1 107 €.

Notre base de données est donc constituée de 7 112 individus qui étaient présents au 1er janvier 2000, avec une légère sur-représentation des femmes (51%). En ce qui concerne le statut d'occupation, 44% des individus travaillent alors que 15% sont retraités, 25% sont entrain de faire des études et 16% sont "inactifs". Cette dernière catégorie regroupe aussi bien les chômeurs que les personnes qui ne recherchent pas d'emploi (hors étudiants et retraités). Les individus en emploi ont une probabilité d'avoir une dépense ambulatoire équivalente aux inactifs (93%) mais le montant de cette consommation ambulatoire est plus faible (952 € vs 1 375 €). Plus on est âgé, plus la probabilité d'avoir une dépense ambulatoire croît. Pour les individus de moins de 30 ans ces deux proportions sont de 89% ; alors qu'elle est de 100 % pour ceux âgés de 75 ans et plus. Il en est de même pour le montant des consommations qui passe de 477 € à 2 937 € entre ces deux tranches d'âges.

La littérature économique de ces deux dernières décennies met clairement en évidence une forte augmentation de la consommation de soin dans la période précédant le décès (time to death). Dans notre panel, nous avons la possibilité d'observer les consommations un an et deux ans avant la mort pour les 345 décédés de notre panel. C'est pour cette variable même si elle concerne assez peu d'individus que nous avons les évolutions les plus fortes surtout en ce qui concerne le montant des dépenses ambulatoires. Ainsi entre les personnes qui ne décèdent pas et celles qui ont une dernière année de vie entre 2000 et 2006 le montant des dépenses ambulatoires est quadruplé (de 1 067 € à 4 201 €). Notre objectif n'est pas d'estimer des causalités mais bien de simuler des trajectoires de consommation sur le cycle de vie dans DESTINIE. La mortalité est simulée dans le modèle en amont de la consommation de soin ambulatoire. Dans une perspective descriptive, il est donc légitime d'intégrer la distance à la date de décès dans le modèle de simulation.

Il existe une très grande hétérogénéité dans les consommations. La distribution des dépenses ambulatoires se révèle être asymétrique vers la gauche. Cette asymétrie se retrouve quelles que soient les caractéristiques socio-économiques étudiées. Cela signifie que le montant des consommations est très important pour un nombre d'individus relativement faible. Ainsi, 55,1 % des hommes et 41,2 % des femmes dépensent moins de 500 € mais 3,1 % des hommes et des femmes dépensent plus de 5 000 € par an et 0,3% des individus ont une consommation supérieure à 15 000 €.

Si nous nous concentrons uniquement sur les individus qui sont vivants en fin d'année 2005 et que l'on regarde leur consommation année après année sur l'ensemble de la période, 76,3 % des individus ont des dépenses ambulatoires toutes les années et pratiquement 9 personnes sur 10 ont consommé au moins pendant 5 années. Très peu d'individus n'ont pas de consommation ambulatoire sur l'ensemble de la période (0.1%). Le montant des dépenses est corrélé à leur fréquence : la médiane du montant des consommations est de 817 € pour ceux qui ont consommé toutes les années contre moins de 200 € pour ceux qui ont consommé moins d'une année sur deux.

### **2.2.1 Les variables retenues pour nos estimations**

Dans nos estimations nous avons retenu:

Table 3: Population

		Nb	$D = 1$	ambulatoire		
				$\bar{m}$ (en €)		
Sexe			Moy	Q1	Q3	
	Homme	3 463	0,91	1 000	123	1 080
	Femme	3 649	0,94	1 207	241	1 491
Statut d'occupation						
	En emploi	3 122	0,93	952	178	1 151
	Retraité	1 131	0,98	2 472	903	2 960
	Inactifs (chômeurs, )	1 096	0,93	1 375	249	1 647
	Etudiant	1 763	0,89	422	81	544
Niveau d'études						
	Autre	595	0,94	1 091	187	1 317
	Primaire	1 822	0,94	1 234	171	1 482
	CAP-BEP	2 489	0,92	1 071	162	1 231
	Bac	933	0,91	1 092	168	1 327
	Supérieur	1 273	0,92	1 016	180	1 221
Etat matrimonial						
	veuf, divorcé,...	711	0,96	2 062	559	2 526
	Marié, ménage	3 747	0,94	1 303	260	1 569
	Célibataire	2 654	0,89	592	93	694
Age						
	- de 30 ans	2 620	0,89	477	86	589
	de 30 à - de 45 ans	1 892	0,93	810	159	985
	de 45 à - de 60 ans	1 307	0,94	1 370	300	1 622
	de 60 à - de 75 ans	947	0,97	2 083	713	2 541
	75 ans et +	346	1,00	2 937	1 232	3 483
Statut par rapport à la mort						
	toujours vivant		0,93	1 067	169	1 283
	mort dans deux ans		0,94	3 417	636	4 405
	mort dans 1 an		0,96	4 201	685	5 083
Ensemble		7 112	0,93	1 107	171	1 308

Table 4: Stat ambulatoire (en euro)

	Hommes		Femmes		Ensemble	
	en %	$\bar{m}$	en %	$\bar{m}$	en %	$\bar{m}$
Pas de dépense	8,9	0	5,7	0	7,3	0
Moins de 500 €	46,2	207	35,5	231	40,7	218
De 500 € à 1 000 €	18	724	20,8	730	19,5	728
De 1 000 € à 5 000 €	23,8	2 044	34,9	2 022	29,5	2 031
De 5 000 € à 15 000 €	2,7	7 743	2,8	7 460	2,7	7 594
15 000 € et plus	0,4	21 635	0,3	20 468	0,3	21 110
Ensemble	100	1 000	100	1 206	100	1 107

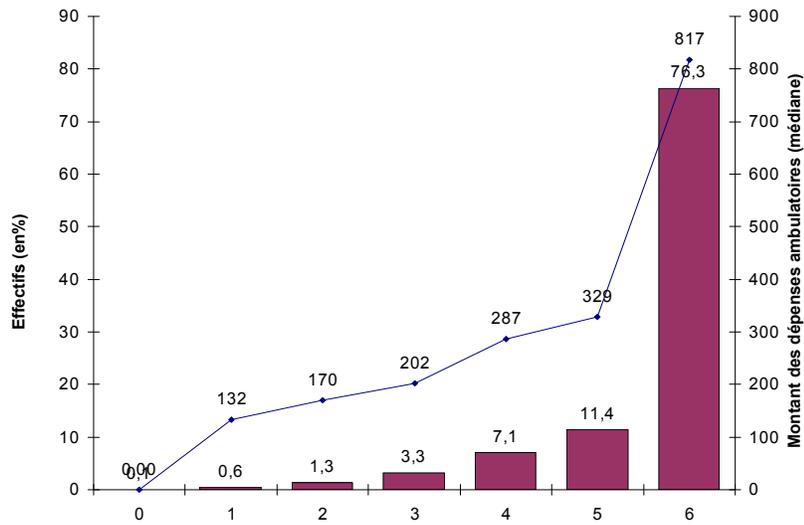


Figure 1: Lien entre fréquence des recours et montants consommés

- AGE: un polynôme d'ordre 4 pour l'âge. Cette dimension qui peut paraître élevée nous permet de tenir compte des différentes inflexions
- TIME TO DEATH: qui nous permet de tenir compte de l'accélération des dépenses avant la mort. Ces dummies nous permettent de savoir si les gens vont mourir dans un an (ttd1) ou dans deux (ttd2). La variable de référence est ne pas décéder dans un futur proche. Nous croisons ces deux variables avec l'âge. Cette importante corrélation entre augmentation des dépenses de santé et fin de vie a été récemment discuté sur données françaises par Brigitte Dormont (Dormont, 2009).
- L'activité professionnelle: cette variable ne concerne que les adultes les modalités sont indépendants, cadres du privé, fonctionnaires, chômeurs, retraités, autres inactifs. La modalité de référence est salariés (non cadre) du privé.
- Le niveau d'étude: Il correspond au niveau d'étude le plus élevé obtenu: primaire, premier cycle, supérieur, autres scolarités. La référence est second cycle.
- Jeune: c'est une dummy qui nous sert à tenir compte du fait qu'un enfant d'un ménage habite toujours chez ses parents. Dès lors pour cet enfant on prend en considération l'activité de la personne de référence du ménage selon les mêmes modalités que précédemment citées. Concernant son niveau d'étude on retient son niveau d'étude actuel selon les mêmes modalités que précédemment citées.

### 3 La modélisation des dépenses de santé

Les distributions des dépenses de santé sont caractérisées par la présence d'un point d'accumulation en zéro (en France:85% des individus pour l'hôpital, 7% pour l'ambulatoire). Les économistes de la santé ont adopté plusieurs stratégies face à cette accumulation de non consommation.

La première stratégie consiste à ignorer le point d'accumulation et à raisonner comme en présence d'une variable continue en effectuant une régression linéaire. Cette stratégie qui a l'avantage de la simplicité a surtout été développée pour expliquer non pas la consommation  $M$  mais la quantité  $\log(M + 1)$ . Elle présente un défaut majeur, elle ne permet pas de modéliser le comportement de recours ou non à une consommation. En effet, la quantité  $P(M = 0|X)$  est estimée à 0 dans un tel modèle. Par ailleurs, d'après Duan (1983), les propriétés des estimateurs obtenus par cette méthode sont plutôt médiocres.

Une autre stratégie consiste à considérer qu'on est en présence d'une variable censurée en 0. La variable continue  $M$  vérifie est une fonction des covariables  $X^M$  :

$$M^* = X^M \beta + \eta \quad (1)$$

$$M = \max(0, M^*) \quad (2)$$

Ce type de modélisation correspond au Tobit de type  $I$  dans la classification proposée par Amemiya en 1985. Un tel modèle suppose non seulement que les déterminants de la non dépense en soins et les déterminants du montant de cette dépense sont identiques, mais elle suppose en plus que les coefficients attachés à ces déterminants sont les mêmes. Cette hypothèse semble peut crédible dans le cas des dépenses de soin.

La troisième façon d'aborder le problème des zéros est de se placer dans une modélisation classique face à des données tronquées et écrire un processus de décision joint (ou sample selection model SSM ou modèle de type *II* dans la classification d'Amemiya (1985)). On écrit alors deux équations distinctes : la première détermine si l'individu va avoir une consommation positive, la seconde détermine son montant. On note  $D_{it}$  le fait que l'individu  $i$  recourt à la date  $t$  ( $D_{it} = 1$  en cas de recours et  $D_{it} = 0$  sinon). On note  $M_{it}$  le montant ou le nombre de jours consommés. Dans les cas où il n'y aura pas d'ambiguïté, on omettra les indices  $i$  et  $t$ . On note  $X^D$  l'ensemble des covariables relatives à  $D$  et  $X^M$  celles relatives à  $M$ . L'ensemble des covariables  $X^D$  et  $X^M$  est noté  $X$ . Nous ne faisons pas a priori l'hypothèse que  $X^D \neq X^M$  mais cette question sera évoqué par la suite. La dépense se déduit de la donnée  $(D_{it}, M_{it})$  :

$$D = \mathbf{1}_{\{X^D\gamma + \varepsilon > 0\}} \Rightarrow P(D = 1|X) = F_{-\varepsilon}(X^D\gamma) \quad (3)$$

$$M = (X^M\beta + \eta)D \quad (4)$$

Afin de prendre en compte une éventuelle corrélation entre les inobservées  $\varepsilon$  qui influent sur la décision de recours et les inobservées  $\eta$  qui influent sur le montant consommé, l'hypothèse suivante est classiquement retenue :

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} | X \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right) \quad (5)$$

Il existe une quatrième manière de modélisée les consommations de soin, connue dans la littérature en économie de la santé sous le nom de Two Part Model (2PM). Le montant des soins consommés est modélisé conditionnellement à la décision de recourir à des soins. L'hypothèse (5) du SSM est en effet remplacé par :

$$E(\eta|D = 1, X) = 0$$

C'est un modèle fondamentalement différent d'un modèle de sélection car dans le Two Part Model la deuxième équation ne prétend modéliser que  $M|X, D = 1$  et non l'ensemble d'une distribution latente  $M|X$  qui ne serait observée que dans le cas  $D = 1$ . Cela signifie que dans le 2PM, on suppose que  $E(\eta|D = 1, X) = 0$ , à la différence du SSM où on suppose que  $E(\eta|D = 1, X) = E(\eta|X\gamma + \varepsilon > 0, X) = \lambda(X) \neq 0$ . Les coefficients  $\beta$  du SSM et du 2PM ne sont pas comparables puisque ce ne sont pas les deux mêmes modèles qui sont estimés. Dans le 2PM, les coefficients  $\beta$  ne peuvent pas s'interpréter comme des impacts causaux des  $X$  sur le montant  $M$  (puisque des inobservées influent sur la sélection  $D$ ). Malgré cette différence d'interprétation (qui joue essentiellement sur l'interprétation économique des  $\beta$ ), on peut alors identifier, estimer et simuler la loi conditionnelle du couple  $(D, M)|X$  grâce à un 2PM car:

$$\mathcal{L}_{D,M|X}(d, m) = [\mathbf{1}_{\{m=0\}} + \mathcal{L}_{M|X,D=1}(m)\mathbf{1}_{\{m>0\}}] P(D = d|X)$$

Les façons d'aborder la modélisation des dépenses en soins dans ces deux modèles est radicalement opposée et a fait l'objet de vifs débats entre les chercheurs de la RAND (Duan *et al.*, 1983; Manning, Duan and Rogers, 1987) partisans du 2PM et d'autres

chercheurs (Hay, Leu and Rohrer, 1987; Hay and Olsen, 1984; Maddala, 1985) partisans d’une modélisation de type SSM. En résumé, les partisans du 2PM reprochent au SSM de nécessiter des hypothèses en terme de distribution (hypothèse 5) dont on ne peut tester la validité (Duan *et al.*, 1983). On peut ajouter que ces hypothèses distributionnelles sont particulièrement sensibles dans le cas où les covariables qui jouent sur  $D$  et  $M$  sont les mêmes ; les SSM sont souvent considérés comme peu robuste en l’absence de variables d’exclusion<sup>1</sup>. Les défenseurs du SSM s’interrogent sur l’existence d’une distribution bi-variée de  $(\varepsilon, \eta)$  qui serait telle que la variable  $(\eta/\varepsilon > -X^D\gamma)$  suive une loi normale (Hay et Olsen, 1984). Ils soulignent que si les termes d’hétérogénéité des équations (3) et (4) ne sont pas indépendants (ce qui est probablement le cas), la distribution de  $(\eta/\varepsilon > -X^D\gamma)$  est fonction de  $X^D\gamma$  et qu’il n’y a alors aucune raison pour que  $E(\eta/\varepsilon > -X^D\gamma) = 0$ . Duan *et al.* (1983) fournissent un contre exemple à ce dernier argument ; ils trouvent une distribution autorisant une corrélation entre  $\varepsilon$  et  $\eta$ , mais qui est telle que la loi de  $\eta$  sur le sous échantillon des personnes ayant recours à des soins suivent une loi normale et soit indépendant des explicatives communes aux deux équations.

Maddala (1985) entre le débat et prend parti pour le SSM considérant qu’il représente mieux le processus de décision qui va engendrer une dépense en soins. Les partisans du 2PM arguant d’un processus séquentiel de décision pour motiver les dépenses, Maddala contre cet argument en disant que s’il y a des variables omises communes aux deux équations (comme l’état de santé par exemple), les termes d’hétérogénéité sont forcément corrélés. Enfin, il ne se dit pas convaincu par le contre exemple fourni par Duan qu’il qualifie de ”sémantique”. Duan *et al.* (1985) répondent à ces arguments en pointant le fait que leurs travaux ne visent pas à estimer les paramètres des équations  $\beta$  mais à prédire le mieux possible les dépenses moyennes; c’est-à-dire estimer  $E(M|X)$  car :

$$\begin{aligned} E(M|X) &= P(D = 1|X) E(M|D = 1, X) \\ &= P(D = 1|X^D) E(M|D = 1, X^M) \end{aligned}$$

Les simulations semblent montrer que lorsque l’objet est de prédire au mieux les dépenses, le 2PM est robuste alors que le SSM peut souffrir de problèmes de multicollinéarité (Manning, Duan and Rogers, 1987; Leung and Yu, 1996). Là encore cependant, ce résultat ne fait pas consensus dans la communauté des chercheurs. Ce résultat est cohérent avec le fait que l’estimation de  $E(M|X, D = 1)$  ne dépend pas de la spécification paramétrique de la loi jointe de  $(\varepsilon, \eta)$  dans le cas d’un 2PM alors que cette spécification paramétrique est indispensable pour un SSM en l’absence d’un instrument à large support<sup>2</sup> (Manski, 1977 ; 1985 ; 1986; 2003).

L’utilisation de modèles de sélection s’impose moins dans le cadre des modèles sur les consommations en soins que dans le cadre des problématiques d’offre de travail. Dans ce dernier cadre, rappelons le raisonnement qui prévaut pour étudier les rendements de l’éducation : toute personne a, en fonction de son capital humain et de ses qualités propres, un salaire potentiel si elle se porte sur le marché du travail. Cependant, en fonction de

<sup>1</sup>On parle de variable d’exclusion à propos d’une variable  $Z$  telle que  $Z \in X^D$  et  $Z \notin X^M$

<sup>2</sup>On parle d’instrument à large support à propos d’une variable  $Z$  telle que  $Z \in X^D$  et  $Z \notin X^M$  et telle que  $P(D = 1|Z, X^D)$  décrit l’ensemble du segment  $[0, 1]$  lorsque  $Z$  varie sur son support. Un instrument à large support est donc une variable d’exclusion particulière.

ce salaire potentiel et des paramètres de leur fonction d'utilité, certaines personnes vont décider de ne pas participer au marché du travail et on ne va donc pas pouvoir observer leur salaire (le salaire offert est inférieur au salaire de réserve). Si on a estimé le rendement de l'éducation sur la productivité (ou sur la capacité à obtenir un salaire élevé), il est important de ne pas raisonner sur les seules personnes pour qui travaillent effectivement mais d'inférer le modèle sur l'ensemble de la population. Il faut notamment prendre en compte qu'à diplôme donné, les personnes qui participent au marché du travail, sont celles qui ont par ailleurs des qualités propres telles qu'elles vont pouvoir négocier un salaire en moyenne plus élevé que les personnes qui ont le même diplôme qu'elles. Dans le cas contraire, on surestime les rendements de l'éducation sur la productivité. L'enjeu est alors d'estimer une covariation du salaire et du niveau d'étude, toute choses observées *et inobservées* égales par ailleurs.

Dans le cas des dépenses de santé, si on souhaite estimer l'influence d'une variable  $X$  toutes choses égales par ailleurs ( $y$  compris l'état de santé qui est inobservable) sur le montant des dépenses, le recours au SSM est indispensable. Il semble cependant que l'ambition de nombreux économètres de la santé n'est pas d'obtenir une telle estimation structurelle mais plutôt d'être capable de faire des prévisions de dépense en fonction de variables observables comme l'âge. Dans ce dernier cas, la covariation de l'âge et des dépenses de santé ne s'interprète pas *stricto sensus* comme un impact causal, mais comme un élément descriptif rendant compte d'une dégradation de l'état de santé avec le vieillissement. C'est l'argument mis en avant par Manning, Duan and Rogers en 1987: lorsqu'on veut prédire les dépenses le fait que les coefficients attachés aux explicatives ne correspondent pas aux impacts causaux n'a pas d'importance.

### 3.1 Le problème de retransformation

Le deuxième grand sujet de la littérature sur les méthodes économétriques en économie de la santé tourne autour du "retransformation problem"<sup>3</sup> (Manning, 1998, Mullahy, 1998; Manning and Mullahy, 2001; ). Dans l'équation (4), la variable  $M$  peut représenter aussi bien le montant des consommations que n'importe quelle transformation injective de ceux-ci. La plupart des économètres de la santé utilisent comme variable  $m$  soit les montants, soit les logarithmes de ceux-ci. Dans le cadre d'un modèle linéaire, travailler sur l'une ou sur l'autre ne repose pas sur les mêmes hypothèses sous-jacentes. Dans le cas où on travaille directement sur les montants, on suppose que les différents facteurs observés  $X^M$  et inobservés  $\eta$  influençant la dépense se combinent additivement; dans le cas où on travaille sur le logarithme des montants, la combinaison des facteurs sur la dépense est multiplicative. Plus généralement, on peut imaginer n'importe quelle transformation de Box-Cox  $x \rightarrow \frac{(x-1)^\lambda}{\lambda}$  pour  $\lambda \geq 0$ . Le cas de la transformation logarithmique correspond à  $\lambda = 0$  et celui d'une modélisation directe des montants correspond à  $\lambda = 1$  ; le cas  $\lambda = \frac{1}{2}$  correspond à une combinaison "quadratique" des facteurs ; le cas  $\lambda = \frac{1}{3}$  correspond à une combinaison "cubique", etc. L'existence d'une queue de distribution très épaisse liée à l'existence de très gros montants de consommation de la part d'une petite partie de la population fait pencher pour un modèle où la combinaison des facteurs est

<sup>3</sup>Par simplicité, les modèles considérés pour les variables continues seront linéaires. Il est a priori possible de considérer des modèles non linéaires mais à ce moment le nombre de covariables doit rester très faible pour ne pas être confronté à des problèmes pratiques liés à l'explosion du nombre d'observations nécessaires pour obtenir des estimations précises. Il ne s'agit plus d'estimer un vecteur dans  $\mathbb{R}^k$  mais une fonction appartenant à un espace fonctionnel beaucoup plus "gros" (problème du "fléau de la dimension").

plutôt multiplicative qu'additive ( $\lambda$  est proche de 0 et non de 1). Les gros consommateurs présenteraient plusieurs facteurs observables (âge) et inobservables (maladie, conduite "à risque") défavorables qui cumulés aurait une résultante plus forte que ce que chaque facteur laisserait penser en le considérant isolément.

Par la suite nous considérerons donc un modèle où les facteurs se combinent de manière multiplicative, nous travaillerons donc non pas sur les montants mais sur leur logarithme<sup>4</sup>.

La transformation logarithmique pose alors un problème : on estime l'espérance conditionnelle de  $E(\ln(M_{it})|D = 1, X^M)$  (ou plus généralement la loi conditionnelle  $\mathcal{L}_{\ln(M_{it})|D=1, X^M}$ ), mais les quantités d'intérêt sont  $E(M_{it}|D = 1, X^M)$  (respectivement  $\mathcal{L}_{M_{it}|D=1, X^M}$ ).

Sous des hypothèses techniques concernant la loi de  $\eta|X^M, D = 1$ <sup>5</sup>, on a :

$$\ln(M) = (X^M \beta + \eta)D \Rightarrow$$

$$\begin{aligned} E(M|X^M, D = 1) &= e^{X^M \beta} \int e^x dF_{\eta|X^M, D=1}(x) \\ &= e^{X^M \beta} \left( e^{E(\eta|X^M, D=1)} + \sum_{k=1}^{+\infty} \frac{E\left[(\eta - E(\eta|X^M, D = 1))^k | X^M, D = 1\right]}{k!} \right) \end{aligned}$$

Rappelons que dans le cas d'un 2PM, on a l'hypothèse de "mean independance" suivante :  $E(\eta|X^M, D = 1) = 0$ . Si on suppose de plus que  $\eta_{it}|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$ , alors  $\int e^x dF_{\eta|X^M, D=1}(x)$  est égale à  $e^{\frac{1}{2}\sigma^2}$ . Plus généralement, en maintenant l'hypothèse de "mean independance"  $E(\eta|X^M, D = 1) = 0$ , on peut relacher l'hypothèse de normalité conditionnelle  $\eta_{it}|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$  et même l'hypothèse d'indépendance conditionnelle :  $\eta \perp\!\!\!\perp X^M | D = 1$  pour estimer  $\beta$  de manière convergente. En revanche, pour estimer  $E(M|X^M, D = 1)$ , il est nécessaire d'estimer  $\int e^x dF_{\eta|X^M, D=1}(x)$  qui dépend de la distribution de  $\eta$  conditionnellement à  $X^M$  et  $D = 1$   $F_{\eta|X^M, D=1}$  et non simplement de l'espérance conditionnelle  $E(\eta|X^M, D = 1)$ . En particulier, si on suppose à tort que  $\eta_{it}|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$ ,  $e^{\frac{1}{2}\sigma^2}$  est un estimateur biaisé de  $\int e^x dF_{\eta|X^M, D=1}(x)$ . Ainsi Duan (1993) propose alors d'utiliser la contrepartie empirique suivante pour estimer  $\int e^x dF_{\eta|X^M, D=1}(x)$  :

$$\hat{E}(M|D = 1, X^M) = e^{X^M \hat{\beta}} \frac{1}{N} \sum_{i=1}^N e^{\hat{\eta}_i}$$

L'estimateur  $\frac{1}{N} \sum_{i=1}^N e^{\eta_i}$  converge vers  $\int e^x dF_{\eta|X^M, D=1}(x) = E(e^\eta) = E(e^\eta|X)$  dès lors que  $\eta \perp\!\!\!\perp X^M | D = 1$ , la normalité conditionnelle n'est donc pas nécessaire. Cependant, comme le montre l'équation 6, si les moments centrés de  $\eta_{it}|X^M, D = 1$  dépendent de  $X^M$ , alors l'estimateur est biaisé (car dans ce cas  $E(e^\eta) \neq E(e^\eta|X)$ ). L'estimateur de Duan est

<sup>4</sup>Mais l'essentiel de ce que nous dirons par la suite resterait valable pour tout autre transformation non linéaire.

<sup>5</sup>Par exemple :  $\eta \perp\!\!\!\perp X^M | D = 1$  et la suite  $\frac{E[(\eta - E(\eta|D=1))^n | D=1]}{n!}$  est bornée.

donc biaisé en présence d'hétéroscédasticité ( $V(\eta_{it}|X^M, D=1)$  dépend de  $X^M$ ). Pour cette raison, un certain nombre de chercheurs à déconseiller de travailler en log.

Face à ce problème de retransformaion, Mullahy (1998) propose d'estimer la deuxième équation du 2PM avec un modèle linéaire généralisé, c'est-à-dire de poser comme hypothèse :  $E(M|D=1, X^M) = \exp\{X^M\beta\}$ .

Dans leur article "Estimating log models : to transform or not transform ?" Manning et Mullahy proposent une série de tests pour décider quelle modélisation est la plus adaptée aux données. En résumé, si la distribution est à queue épaisse, ils recommandent de travailler sur le logarithme des dépenses ; dans le cas contraire, ils préconisent plutôt un modèle linéaire généralisé pour éviter le problème du passage à l'exponentielle.

Une autre alternative consiste à modéliser explicitement les moments de  $\eta$ , par exemple en supposant que :  $\eta_{it}|X^M, D=1 \sim \mathcal{N}(0, e^{X^M\lambda})$ . Cette approche présente les avantages et les inconvénients des méthodes paramétriques : elle est simple à mettre en oeuvre, elle permet de faire des simulations facilement mais elle nécessite des hypothèses paramétriques fortes.

Nous allons voir maintenant que prendre en compte l'hétérogénéité individuelle et la dépendance d'état sur des données longitudinales nous oblige à recourir à un SSM plutôt qu'à un TPM, et à utiliser à des méthodes paramétriques. Nous avons donc décidé de traiter l'hétéroscédasticité par une modélisation paramétrique. A partir de maintenant et dans le reste de cet article, nous travaillerons donc en log.

## 4 Estimation sur données longitudinale

### 4.1 Les gens ne sont pas tous les mêmes

Une importante hétérogénéité individuelle caractérise la plupart des comportements de consommation. Le fait de travailler sur des données en panel permet de mieux prendre en compte cette hétérogénéité. C'est d'ailleurs un enjeu important si on veut décrire, identifier, estimer et à plus forte raison simuler des trajectoires. En effet, des différences durables de comportements entre différents individus laissent penser qu'il existe des caractéristiques inobservées des individus constantes au cours du temps.

Supposons que conditionnellement aux observables  $X$ , on ait alors le modèle suivant :

$$\begin{aligned} D_{it} &= \mathbf{1}_{\{X_{it}^D\gamma + u_i + \varepsilon_{it} > 0\}} \\ \ln(M_{it}) &= (X_{it}^M\beta + v_i + \eta_{it})D_{it} \end{aligned}$$

Les quantités  $u_i$  et  $v_i$  s'interprètent de la manière suivante : certains individus, à caractéristiques  $X$  identiques ont plus souvent recours à des soins que d'autres, et parmi les individus ayant un recours à des soins certains consomment plus que d'autres (à caractéristiques  $X$  identiques), mais cette différence entre individus n'est pas quelque chose de directement observable ; elle ne peut que s'estimer à partir de la fréquence des recours et des montants moyens observés par individu. Cela permet de considérer

des corrélations entre des évènements survenant à des dates différentes : on relâche l'hypothèse  $(D_{it}, M_{it}) \perp\!\!\!\perp (D_{it'}, M_{it'})|X$  qui est irréaliste. Il est d'autant plus indispensable de relâcher cette hypothèse que notre but final est de simuler des trajectoires cohérentes inter-temporellement afin de faire des bilans sur le cycle de vie.

On suppose que les couples  $(u_i, v_i)$  sont indépendants et identiquement distribués. On suppose de plus qu'ils sont indépendants des  $X$ , on recourt d'onc à un modèle de panel à "effet aléatoire". Cela n'est théoriquement pas nécessaire : on pourrait très bien recourir à un modèle de panel à "effets fixes" sans faire d'hypothèse sur la distribution jointe  $(u_i, v_i, X)$ . Cela poserait néanmoins des difficultés d'estimation lié au problème des paramètres incidents. Kyriazidou (1997) propose une méthode d'estimation asymptotiquement sans biais des paramètres  $\beta$  et  $\gamma$  en utilisant les observations telles que  $\gamma_{it}X_{it}^D = \gamma_{it'}X_{it'}^D$ , cette stratégie n'est pas forcément légitime en présence de forts effets d'âge. De plus l'hétérogénéité  $u_i, v_i$  est traitée comme un paramètre de nuisance mais l'estimation de sa distribution n'est pas envisagée, alors que dans une optique descriptive c'est un élément important. Notre ambition est de faire des simulations sur le cycle de vie, donc de simuler aussi des quantités  $u_i$  et  $v_i$ . Si on estimait un modèle à effets fixes, il faudrait en plus estimer les corrélations entre  $u_i, v_i$  et les  $X$  pour faire une imputation cohérente. En recourant à un modèle à effet aléatoire on effectue directement cette projection, mais les coefficients  $\gamma$  et  $\beta$  ne peuvent être interprétés comme l'impact causal des  $X$  si certaines caractéristiques inobservées constantes au cours du temps sont simultanément corrélés aux  $X$  et à  $(D, \ln(M))$ . Encore une fois, dans une optique descriptive afin de faire de la micro-simulation sur le cycle de vie, ce n'est pas forcément gênant. Enfin, à notre connaissance il n'y a pas eu d'applications empiriques de la méthode de Kyriazidou, les difficultés empiriques sont donc relativement peu documentées. Nous réservons donc l'implémentation de cette stratégie d'estimation pour des travaux futurs.

Comme dans un modèle en coupe on suppose que les couples  $(\varepsilon_{it}, \eta_{it})$  sont indépendants et identiquement distribués et indépendants des  $X$ . Enfin on fait comme hypothèse que  $(u_i, v_i) \perp\!\!\!\perp (\varepsilon_{it}, \eta_{it})|X$ .

## 4.2 L'hétérogénéité individuelle invite à privilégier le SSM

Reste à savoir si on fait des hypothèses de type 2PM  $E(v_i + \eta_{it}|X_{it}^M, D_{it} = 1) = 0$  ou de type SSM (dépendance des variables  $u_i, v_i, \varepsilon_{it}, \eta_{it}$ ). Comme  $E(v_i + \eta_{it}|X_{it}^M, D_{it} = 1) = E(v_i|X_{it}^M, u_i > -X_{it}^D\gamma - \varepsilon_{it}) + E(\eta_{it}|X_{it}^M, \varepsilon_{it} > -X_{it}^D\gamma - u_i)$ , il s'agit donc préciser les hypothèses que l'on souhaite faire sur les lois jointes de  $u$  et  $v$  d'une part et sur celle de  $\varepsilon$  et  $\eta$  d'autre part. La discussion de la pertinence de l'hypothèse  $\varepsilon_{it} \perp\!\!\!\perp \eta_{it}|X$  d'une part et de l'hypothèse  $\varepsilon_{it}, \eta_{it}|X \sim \mathcal{N}\left((0, 0), \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right)$  a été menée en 3.

L'hypothèse  $u_i \perp\!\!\!\perp v_i|X$  implique que le montant consommé à une date est indépendant (conditionnellement aux  $X$ ) du nombre de recours sur l'ensemble des périodes observées. Ce résultat est assez intuitif puisque les variables inobservées  $u_i$  impliquent une plus ou moins grande fréquence de recours, alors que les variable inobservées  $v_i$  implique un plus ou moins grand montant moyen de consommation. La preuve formelle du résultat est renvoyée en annexe. Cette hypothèse est clairement rejetée par les données, les personnes qui recourent souvent à des soins une année donnée ont des montants de consommation

annuelle nettement plus élevés que les autres. On a donc corrélation entre  $u_i$  et  $v_i$  (conditionnellement aux  $X$ ), on est donc naturellement conduit à privilégier un SSM sur un 2PM dans le cadre d'une estimation longitudinale.

Dans le cas des modèles de sélection et en l'absence d'instrument à large support, il est usuel (et nécessaire) de faire des hypothèses paramétriques concernant la distribution des variables inobservées. Pour estimer le modèle on suppose<sup>6</sup> donc maintenant que :

$$(u, v, \varepsilon, \eta) \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_1 \sigma_u \sigma_v & 0 & 0 \\ \rho_1 \sigma_u \sigma_v & \sigma_v^2 & 0 & 0 \\ 0 & 0 & 1 & \rho_2 \sigma_\eta \\ 0 & 0 & \rho_2 \sigma_\eta & \sigma_\eta^2 \end{bmatrix} \right)$$

### 4.3 La dépendance d'état et le problème des conditions initiales

Les intuitions que l'on peut avoir sur la dynamique de l'état de santé ainsi que le modèle théorique de Grossman implique une dépendance d'état sur le consommation de soin. Bien que de nombreux auteurs disposent de données de panel permettant l'introduction d'une dynamique dans les dépenses en soins, on trouve assez peu d'exemples de modélisations dynamiques en économie de la santé (Jones and Rice, 2004; Bago d'Uva, 2005; Nolan, 2007).

Formellement, le modèle envisagé s'écrit de la manière suivante pour la date  $t > 0$  :

$$D_{it} = \mathbf{1}_{\{D_{it-1}\gamma_1 + \ln(M_{it-1})\gamma_2 + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}} \quad (6)$$

$$\ln(M_{it}) = (D_{it-1}\beta_1 + \ln(M_{it-1})\beta_2 + X_{it}\beta + v_i + \eta_{it})D_{it} \quad (7)$$

Si on veut avoir une interprétation structurelle du modèle, se pose alors un problème de taille pour l'estimation des coefficients  $\gamma_1, \gamma_2, \gamma, \beta_1, \beta_2, \beta$  car par nature, le couple  $D_{it-1}, \ln(M_{it-1})$  est corrélé au couple  $u_i, v_i$ . Cela est-il gênant si on veut obtenir une "bonne" description de la trajectoire  $D_{it}, M_{it}$  afin de la simuler. Raisonner comme on le fait pour les covariables  $X$  ne permet de surmonter le problème. En effet, pour faire une "bonne" simulation, il faudra simuler conjointement  $(u_i, v_i)$  et  $D_{it-1}, M_{it-1}$ . En appliquant des conditionnements successifs sur les dates antérieures, il apparaît que l'enjeu est de simuler conjointement l'hétérogénéité individuelle  $(u_i, v_i)$  et les conditions initiales  $D_{i0}, \ln(M_{i0})$ .

En effet, la loi de la trajectoire se décompose de la manière suivante :

$$\mathcal{L}_{(D_t, \ln(M_t))_{t=0 \dots T} | X} = \int_u \int_v \prod_{t=1}^T \mathcal{L}_{D_t, \ln(M_t) | X, u, v, D_{t-1}, \ln(M_{t-1})} \mathcal{L}_{D_0, \ln(M_0), u, v | X} du dv$$

Mais les équations (6) et (7) implique que  $(D_0, \ln(M_0))$  est corrélé avec  $(u, v)$ . On ne peut donc décomposer la loi  $\mathcal{L}_{D_0, \ln(M_0), u, v | X}$  en un produit de loi marginale :

$$\mathcal{L}_{D_0, \ln(M_0), u, v | X} \neq \mathcal{L}_{D_0, \ln(M_0) | X} \mathcal{L}_{u, v | X}$$

Estimer un tel modèle dynamique est donc difficile dans la mesure où la loi jointe de  $D_0, \ln(M_0), u, v$  n'est pas facilement identifiable. Et faire l'hypothèse que  $D_0, \ln(M_0)$  est indépendante de  $(u, v)$  est intrinsèquement contradictoire avec le modèle. Ce problème traduit la difficulté à séparer dépendance d'état et hétérogénéité individuelle. Les conditions initiales sont endogènes.

<sup>6</sup>Comme d'habitude, on peut normaliser la variance de  $\varepsilon$  à 1 car le paramètre  $\gamma$  n'est identifiable qu'à une constante multiplicative près.

### 4.3.1 L'approche d'Heckman (1981)

Heckman a étudié le problème d'endogénéité des conditions initiales dans un modèle binaire avec dépendance d'état. Si on extrapole cette méthode au cas qui nous intéresse ici, il s'agit d'approcher la loi de  $D_0, \ln(M_0)|u, v, X$  en postulant que les variables latentes associées en première période vérifient une relation du type :

$$\begin{pmatrix} D_0^* \\ \ln(M_0) \end{pmatrix} | u, v, X \sim \mathcal{N} \left( \begin{bmatrix} X\gamma_0 + \mu_1 u + \mu_2 v \\ X\beta_0 + \mu_3 u + \mu_4 v \end{bmatrix}, \begin{bmatrix} 1 & \rho_0 \sigma_0 \\ \rho_0 \sigma_0 & \sigma_0^2 \end{bmatrix} \right)$$

Cette relation n'est pas structurelle dans le sens où les paramètres  $\gamma_0, \beta_0, \mu_i (i = 1..4)$  n'ont pas de sens économique. De plus, même la forme utilisée pour modéliser la loi de  $D_0, \ln(M_0)|u, v, X$  n'est pas cohérente avec la dynamique (à la différence du modèle linéaire).

En pratique l'estimation est assez difficile à réaliser, il est difficile de "séparer" les  $\beta$ , les  $\gamma$ , les  $\beta_0$  et les  $\gamma_0$ . De plus la convergence des estimateurs n'est assurée que quand  $T \rightarrow +\infty$ . Ainsi dans ses travaux sur le modèle binaire, Heckman met en évidence sur des données simulées que les estimateurs peuvent être fortement biaisés si le nombre de périodes est faible. Nous disposons d'un panel de 6 périodes ce qui rend inapproprié le recours à une telle méthode.

### 4.3.2 L'approche de Wooldridge

Wooldridge (Wooldridge, 2005) prend un parti différent de celui d'Heckman ; au lieu de faire des hypothèses sur la loi conditionnelle de  $D_0, \ln(M_0)|u, v, X_0$ , il propose de faire des hypothèses sur la loi de  $(u, v)|D_0, \ln(M_0), (X_t)_{t=0..T}$ .

$$\begin{aligned} (u_i, v_i) | D_{i0}, \ln(M_{i0}), X_i &\sim \mathcal{N} (D_{i0}K_D + \ln(M_{i0})K_M + X_iK_X, \Sigma) \\ &= \mathcal{N} \left( \begin{bmatrix} \kappa_{uD}D_{i0} + \kappa_{uM}\ln(M_{i0}) + \sum_{t=0}^T X_{it}\kappa_{uX_{it}} \\ \kappa_{vD}D_{i0} + \kappa_{vM}\ln(M_{i0}) + \sum_{t=0}^T X_{it}\kappa_{vX_{it}} \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_1 \sigma_u \sigma_v \\ \rho_1 \sigma_u \sigma_v & \sigma_v^2 \end{bmatrix} \right) \end{aligned}$$

C'est une approche très voisine de celle de Chamberlain pour contourner le problème des paramètres incidents dans le cas des modèles de panel non linéaire. On décompose l'hétérogénéité entre une partie corrélée aux  $X$  (terme du type  $X_iK_X$ ) et une partie non corrélée modélisée par un "effet aléatoire" normal. Ici, il ne s'agit pas de contrôler une éventuelle corrélation entre l'hétérogénéité et les  $X$  mais entre l'hétérogénéité et les conditions initiales.

En remarquant qu'on a alors :

$$\mathcal{L}_{(D_t, \ln(M_t))_{t=0..T} | X} = \int_u \int_v \prod_{t=1}^T \mathcal{L}_{D_t, \ln(M_t) | X, u, v, D_{t-1}, \ln(M_{t-1})} \mathcal{L}_{D_0, \ln(M_0) | X} \mathcal{L}_{u, v | X, D_0, \ln(M_0)} du dv$$

On peut alors estimer le modèle suivant pour  $t > 0$ :

$$\begin{aligned} D_{it} &= \mathbf{1}_{\{D_{it-1}\gamma_1 + \ln(M_{it-1})\gamma_2 + X_{it}\gamma + D_{i0}\kappa_{uD} + \ln(M_{i0})\kappa_{uM} + \sum_{t=0}^T X_{it}\kappa_{uX_{it}} + u_i + \varepsilon_{it} > 0\}} \\ \ln(M_{it}) &= \left( D_{it-1}\beta_1 + \ln(M_{it-1})\beta_2 + X_{it}\beta + D_{i0}\kappa_{vD} + \ln(M_{i0})\kappa_{vM} + \sum_{t=0}^T X_{it}\kappa_{vX_{it}} + v_i + \eta_{it} \right) D_{it} \end{aligned}$$

Et estimer la loi de  $D_0, \ln(M_0)|X$  "en coupe" pour simuler l'initialisation de la trajectoire (terme  $\mathcal{L}_{D_0, \ln(M_0) | X}$ ).

Les termes d'hétérogénéité inobservée constante au cours du temps sont :  $D_{i0}\kappa_{uD} + \ln(M_{i0})\kappa_{uM} + \sum_{t=0}^T X_{it}\kappa_{uX_{it}} + u_i$  et  $D_{i0}\kappa_{vD} + \ln(M_{i0})\kappa_{vM} + \sum_{t=0}^T X_{it}\kappa_{vX_{it}} + v_i$ . Les composantes  $D_{i0}\kappa_{uD} + \ln(M_{i0})\kappa_{uM}$  et  $D_{i0}\kappa_{vD} + \ln(M_{i0})\kappa_{vM}$  permettent de rendre compte de la corrélation entre les conditions initiales et l'hétérogénéité inobservée ; les composantes  $\sum_{t=0}^T X_{it}\kappa_{uX_{it}}$  et  $\sum_{t=0}^T X_{it}\kappa_{vX_{it}}$  rendent compte de la corrélation entre les  $X$  et l'hétérogénéité inobservée ; et enfin les termes  $u_i$  et  $v_i$  représentent les composantes de l'hétérogénéité qui ne sont pas corrélées avec les autres variables.

Un des gros intérêts de la méthode de Wooldridge réside dans la facilité de mise en oeuvre. Les variables  $D_0, \ln(M_0)$  et  $(X_t)_{t=0..T}$  jouent donc exactement le même rôle dans les estimations que les covariables  $X$ , et les techniques d'estimation présentées précédemment s'appliquent (maximum de vraisemblance, estimation en deux étapes et corrections par ratio de Mills etc.).

Autre intérêt, contrairement à la procédure proposée par Heckman, la convergence des estimateurs pour  $T < +\infty$  est assurée (sous réserve que l'hypothèse paramétrique sur la distribution de l'hétérogénéité  $u, v$  soit vérifiée).

Cependant le nombre de paramètres à estimer peut devenir très grand ( $K_X$  est de taille  $K \times T$ ), de plus il peut se poser des problèmes de quasi-colinéarité entre les  $X$  et les  $(X_t)_{t=0..T}$  ainsi qu'entre  $D_{t-1}, \ln(M_{t-1})$  et  $D_0, \ln(M_0)$ . Sur nos données, il s'est révélé impossible de séparer les différents effets de manière convaincante par cette méthode, principalement du fait de la corrélation de l'âge et de la consommation à la date  $t = 0$  avec l'âge et la consommation aux dates suivantes. Nous nous sommes retrouvé confronté à ce problème sur les données que nous avons exploitées, nous proposons donc dans la section suivante une adaptation de la méthode de Wooldridge pour contourner ces difficultés.

### 4.3.3 Une adaptation de la méthode de Wooldridge

Reprenons la philosophie de la méthode de Wooldridge. Il s'agit essentiellement de supposer que l'hétérogénéité est de la forme :

$$(u, v) | D_0, \ln(M_0), X \sim \mathcal{N}(f(D_0, \ln(M_0), X), \Sigma)$$

Où  $f$  appartient à un ensemble de fonction. Dans sa version originale, Wooldridge propose de considérer l'ensemble des fonctions  $f$  linéaires. Cet ensemble de fonction est donc relativement "gros" si il y a beaucoup de variables  $X$  et beaucoup de périodes. L'impossibilité de séparer les effets de manière convaincante sur nos données s'explique donc par une trop grande flexibilité dans la modélisation de l'hétérogénéité inobservée.

Nous nous proposons donc de reformuler l'hypothèse de Wooldridge en considérant un ensemble de fonction  $f$  plus restreint que l'ensemble des fonctions linéaires.  $u, v$  doit représenter la sur-propension (respectivement la sous-propension) à recourir à une consommation et la sur-propension (respectivement la sous-propension) à avoir des dépenses élevées en cas de recours, compte tenu des caractéristiques  $X$  de l'individu. Le problème des conditions initiales est précisément que cette sur ou sous propension à consommer est corrélée à la sur ou sous propension à consommer à la date  $t = 0$ . Nous nous proposons donc d'utiliser le résidu généralisé de  $r_0$  qui est la meilleure estimation possible de la sur ou de la sous propension à consommer à la date initiale. Dans un cadre linéaire le résidu généralisé serait simplement le résidu "classique"  $Y_0 - E[Y_0|X_0]$ . Dans le cadre d'un modèle logit ou probit ( $Y_0 = \mathbf{1}_{\{X_0\delta + \xi\}}$ ) le résidu généralisé serait  $E(\xi|Y_0, X_0)$ .

On est alors ramené à définir le processus générateur de données pour le modèle en coupe à la date d'entrée dans le panel. Nous avons eu recours à un 2PM en  $t = 0$  pour

estimer les résidus généralisés pour la raisons suivantes :

- Un processus générateur de données de type SSM en  $t = 0$  n'est pas compatible avec le modèle dynamique retenu pour  $t > 0$
- L'absence de variable d'exclusion rend les estimations d'un SSM sur la seule première période très instables
- Cette estimation en  $t = 0$  n'est pas structurelle, nous ne voulons pas interpréter les paramètres  $\gamma_0$  et  $\beta_0$  comme des impacts causaux, le recours à un SSM n'est donc pas indispensable. Il s'agit essentiellement d'avoir une approche descriptive pour appréhender la sur ou la sous propension à consommer  $r_{D_0}, r_{ln(M_0)}$  à la date  $t = 0$  d'un individu ayant les caractéristiques  $X_0$ . Dans une perspective descriptive, le 2PM est plus robuste à l'hétéroscédasticité et à la non normalité des résidus sur l'équation des montants

Le résidu généralisé s'écrit<sup>7</sup> donc :

Pour un consommant en t=0 :

$$r_0 = (r_{D_0}, r_{ln(M_0)}) = \left( \frac{\phi(X_0\gamma_0)}{\Phi(X_0\gamma_0)}, ln(M_0) - X_0\beta_0 \right)$$

Pour un non consommant en t=0 :

$$r_0 = (r_{D_0}, r_{ln(M_0)}) = \left( -\frac{\phi(X_0\gamma_0)}{1 - \Phi(X_0\gamma_0)}, 0 \right)$$

On peut alors postuler que l'hétérogénéité individuelle est distribuée selon la loi conditionnelle suivante :

$$(u, v) | D_0, ln(M_0), X \sim \mathcal{N} \left( (r_{D_{i0}}\gamma_{r_{D_0}} + r_{ln(M_{i0})}\gamma_{r_{ln(M_0)}}, r_{D_{i0}}\beta_{r_{D_0}} + r_{ln(M_{i0})}\beta_{r_{ln(M_0)}}), \begin{bmatrix} \sigma_u^2 & \rho_1\sigma_u\sigma_v \\ \rho_1\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix} \right)$$

#### 4.4 Le problème de l'hétéroscédasticité

Le problème des conditions initiales étant réglé, reste à prendre en compte une éventuelle hétéroscédasticité des résidus dans le modèle dynamique. Le problème que pose l'hétéroscédasticité dans un SSM en coupe a été rappelé en 3.1.

<sup>7</sup>Dans notre problème si on pense que le processus générateur de données à la date  $t = 0$  est un SSM, le résidu généralisé s'exprime de la manière suivante :

Pour un consommant en t=0 :

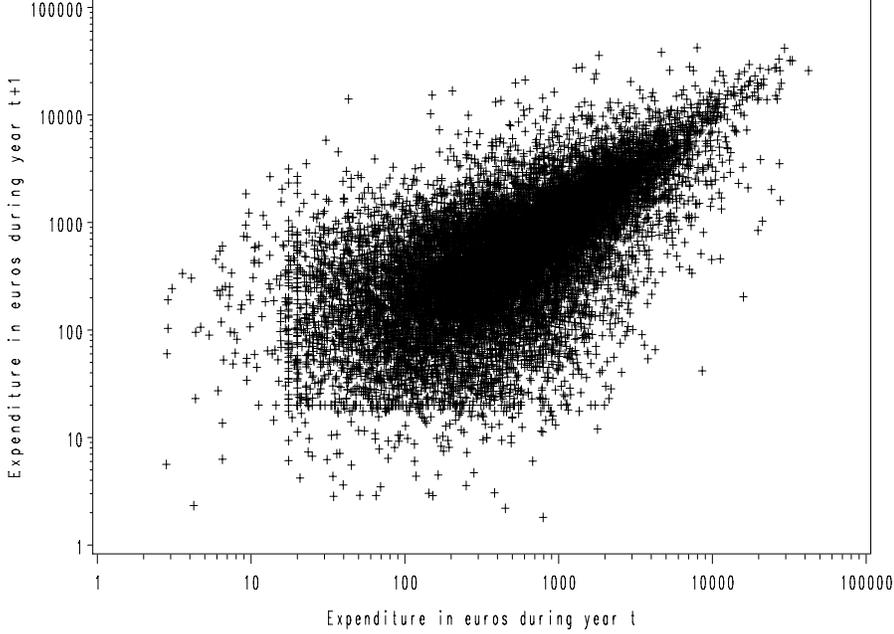
$$r_0 = (r_{D_0}, r_{ln(M_0)}) = \left( \frac{\phi \left( \frac{X_0\gamma_0 + \rho_0 \frac{ln(M_0) - X_0\beta_0}{\sigma_{\eta_0}}}{\sqrt{1 - \rho_0^2}} \right)}{\Phi \left( \frac{X_0\gamma_0 + \rho_0 \frac{ln(M_0) - X_0\beta_0}{\sigma_{\eta_0}}}{\sqrt{1 - \rho_0^2}} \right)}, ln(M_0) - X_0\beta_0 \right)$$

Pour un non consommant en t=0 :

$$r_0 = (r_{D_0}, r_{ln(M_0)}) = \left( -\frac{\phi(X_0\gamma_0)}{1 - \Phi(X_0\gamma_0)}, -\rho_0\sigma_{\eta} \frac{\phi(X_0\gamma_0)}{1 - \Phi(X_0\gamma_0)} \right)$$

Les statistiques descriptives mettent en évidence une forte hétéroscédasticité du logarithme des montants : en effet le nuage de points représentant le log des montants de deux consommations successives est fortement aplati pour les gros montants.

**Ambulatory expenditure for men**



La non prise en compte de cette hétéroscédasticité est susceptible de biaiser l'estimation des coefficients  $\beta$  et  $\gamma$  (contrairement au modèle linéaire), et elle pose de plus des problèmes de retransformation, le niveau moyen des montants étant relié à la dispersion des log-montants (car  $E(e^\eta) = e^{\frac{\sigma^2}{2}}$  dans le cas gaussien). Il est donc nécessaire de modéliser également une éventuelle hétéroscédasticité. C'est ce que nous avons fait en supposant que :

$$\eta|D_{t-1}, m_{t-1}, X_t \sim \mathcal{N}(0, e^{(\lambda_0 + \lambda_D D_{t-1} + \lambda_{\ln(M_{t-1})} \ln(M_{i,t-1}))})$$

On estime pour le modèle avec hétéroscédasticité :

$$D_{it} = \mathbf{1}_{\{D_{it-1}\gamma_1 + \ln(M_{it-1})\gamma_2 + \hat{r}_{D_{i0}}\gamma_{r_{D_0}} + \hat{r}_{\ln(M_{i0})}\gamma_{r_{\ln(M_0)}} + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}}$$

$$\ln(M_{it}) = \left( D_{it-1}\beta_1 + \ln(M_{it-1})\beta_2 + \hat{r}_{D_{i0}}\beta_{r_{D_0}} + \hat{r}_{\ln(M_{i0})}\beta_{r_{\ln(M_0)}} + X_{it}\beta + v_i + \eta_{it} \right) D_{it}$$

La matrice de variance covariance de  $(u_i, v_i, \varepsilon_{it}, \eta_{it})$  est :

$$\begin{bmatrix} \sigma_u^2 & \rho_1 \sigma_u \sigma_v & 0 & 0 \\ \rho_1 \sigma_u \sigma_v & \sigma_v^2 & 0 & 0 \\ 0 & 0 & 1 & \rho_2 e^{(\lambda_0 + \lambda_{D_{t-1}} D_{i,t-1} + \lambda_{\ln(M_{t-1})} \ln(M_{i,t-1}))} \\ 0 & 0 & \rho_2 e^{(\lambda_0 + \lambda_{D_{t-1}} D_{i,t-1} + \lambda_{\ln(M_{t-1})} \ln(M_{i,t-1}))} & e^{2(\lambda_0 + \lambda_{D_{t-1}} D_{i,t-1} + \lambda_{\ln(M_{t-1})} \ln(M_{i,t-1}))} \end{bmatrix}$$

Et  $\hat{r}_{\ln(M_{i0})} = \hat{E}(\tilde{\varepsilon}_{i0}|D_{i0}, \ln(M_{i0}), X_{i0})$  et  $\hat{r}_{D_{i0}} = \hat{E}(\tilde{\eta}_{i0}|D_{i0}, \ln(M_{i0}), X_{i0})$  sont les estimateurs des résidus généralisés du 2PM suivant :

$$D_{i0} = \mathbf{1}_{\{X_{i0}\gamma_0 + \tilde{\varepsilon}_{i0} > 0\}}$$

$$\ln(M_{i0}) = (X_{i0}\beta_0 + \tilde{\eta}_{i0}) D_{i0}$$

## 5 Estimation des dépenses de santé

Nous reprenons ici deux estimations : l'une correspondant à un Two-part model estimé en coupe, l'autre correspondant à un modèle de selection avec dépendance d'état, hétérogénéité inobservée et hétéroscédasticité des résidus sur l'équation de montant. La première estimation permet de simuler des dépenses à une date donnée, mais ne permet pas d'imputer des trajectoires de consommation qui soient cohérentes dans la dimension temporelle. La deuxième estimation prend en compte la dépendance temporelle de la variable à simuler. Les deux estimations permettent d'estimer les quantités  $E\left(\sum_{t=0}^T \frac{M_{it}}{(1+r)^t} | X\right)$ , mais seule la deuxième permet d'estimer de manière crédible les quantités  $E\left(\sum_{t=0}^T \frac{M_{it}}{(1+r)^t} | X\right)$ . Nous estimons ces deux modèles sur les hommes et les femmes séparément.

L'intérêt d'estimer également un two-part model en coupe est que les caractéristiques observables corrélées à une forte consommation de soin apparaissent plus clairement que dans un modèle avec dépendance d'état où elles jouent également sur la variable retardée.

D'autres modèles peuvent être estimés : modèle de panel sans dépendance d'état, sans prise en compte de l'hétéroscédasticité ect. pour une discussion détaillée de ces différents modèles, on pourra se reporter à Albouy, Davezies et Debrand (2009).

### 5.1 Différence entre groupes sociaux

Globalement, à âge comparable, les gros consommateurs sont plutôt des cadres du privé par opposition aux indépendants qui consomment moins. Les chômeurs recourent moins souvent à des soins que les ouvriers et employés mais les montants consommés sont plus élevés en cas de recours. Les inactifs non retraités ont des consommations relatives très différentes chez les hommes et chez les femmes, conséquence du fait que les déterminants de la participation au marché du travail diffèrent selon le genre. Les enfants de cadre ont des consommations plus importantes que les enfants d'employés et d'ouvriers. A contrario, les enfants de chômeurs de retraités et d'inactifs consomment moins. Les enfants d'indépendant recourent moins souvent à une dépense ambulatoire que les enfants de d'employés et d'ouvriers mais en cas de recours leur dépenses sont comparables à celles des enfants d'ouvriers et d'employés.

Cependant, ces différences de consommation en coupe ne permettent pas à elles seules d'estimer des différences de consommation sur le cycle de vie. En effet, les différences d'espérance de vie sont très importantes entre ces différents milieux sociaux (jusqu'à huit ans d'écart entre un cadre et un ouvrier).

Un autre point important concerne la forte corrélation temporelle des dépenses.

### 5.2 Importance de la corrélation dans la dimension longitudinale

La corrélation des dépenses dans la dimension longitudinale est due à deux médiateurs par deux types de paramètres :

- Des paramètres rendant compte d'une hétérogénéité inobservée :  $\sigma_u, \sigma_v$  et paramètres relatifs aux conditions initiales. Tous les estimations de ces paramètres sont relativement élevées et fortement significatives.

- Des paramètres relatifs à une dépendance d'état : effet du recours et de la consommation à la date précédente. Là aussi, les estimations de ces paramètres indiquent une forte dépendance d'état.

La prise en compte de cette corrélation au cours du temps est crucial si on cherche à décrire le risque médical sur le cycle de vie.

## 6 Simulation des dépenses de santé dans Destinie

A compléter en fonction des améliorations à venir de DESTINIE sur la mortalité différentielle

### 6.1 Exemple de programme PERL

Nous donnons ici à titre d'illustration un petit morceau de programme relatif à la simulation du modèle dynamique estimé. Les paramètres sont dans des fichiers excel, ils ont été lus et stockés dans un "hachage". La fonction `simut2`, appelée comme argument un individu (repéré par le scalaire `$` ou `$i`) et le hachage contenant les coefficients (repéré par le hachage `%` ou `%coef`). La fonction `simut2` utilise les fonctions intermédiaires suivantes :

- `simu0` : permet d'imputer une consommation à la naissance, calcule également les résidus généralisés de cette imputation.
- `creavar` : de lire et de stocker l'ensemble des covariables pour un individu
- `Norm2` : permet de générer des lois normales indépendantes

```
sub simut2($%)
{
#Déclaration des variables utiles pour la suite

my $i=shift(@_); #repère l'individu
my %coef=@_; #récupère les coeffs
my $genre; #définit une variable genre
if ($sexe[$i]==1){$genre="homme"};if ($sexe[$i]==2){$genre="femme"};
my $t; #repère l'age
my $k;
my ($ef1,$ef2,$prop,$logmontant,@parti,@logmontanti,@xbeta,@xgamma,$seps,$eta);
#repère les effets fixes ($ef1 et $ef2), la trajectoire des recours (@parti),
# celle des log-montants consommés (@logmontanti), la trajectoire des
# propensions (@xbeta et @xgamma), les variables $k,$prop,$logmontant,$seps,$eta
# ne sont que des intermédiaires de calcul.
my %var;

my ($part0,$logmontant0,$rd0,$rm0,$xgamma0,$xbeta0)=simu0($i,%coef);
# La fonction simu0 permet de simuler les consommations à la naissance

#Tirage des effets fixes
```

```

$ef1=$coef{$genre}{"dyn_disp"}{"sigma_u"}*Norm2;
if ($coef{$genre}{"dyn_disp"}{"sigma_u"} ne 0)
{
  $ef2=$coef{$genre}{"dyn_disp"}{"sigma_v"}*
    ( $coef{$genre}{"dyn_disp"}{"rho1"}*$ef1/$coef{$genre}{"dyn_disp"}{"sigma_u"}+
    sqrt(1-$coef{$genre}{"dyn_disp"}{"rho1"}**2)*Norm2 );
}
if ($coef{$genre}{"dyn_disp"}{"sigma_u"}==0)
{
  $ef2=$coef{$genre}{"dyn_disp"}{"sigma_v"}*Norm2;
}

#Imputation de la trajectoire date par date
for $t (0..$amort[$i]-$anaiss[$i])
{
#La première date est traitée à part
$prop=0;$logmontant=0;
if ($t==0)
{
  $parti[0]=$part0;
  $logmontanti[0]=$logmontant0;
  $xgamma[0]=$xgamma0;
  $xbeta[0]=$xbeta0;
}

if ($t>0)
{
  %var=creavar($i,$t);
#Ajout de rd0 et rm0 et des retardés au hachage %var
  $var{"rd0"}=$rd0;
  $var{"rm0"}=$rm0;
  $var{"lag_bin"}=$parti[$t-1];
  $var{"lag_log"}=$logmontanti[$t-1];

#Calcul des prediceurs
  foreach $k (keys %{$coef{$genre}{"dyn_d"}})
  {
    if ($k ne "constante")
    {
      $prop=$prop+$coef{$genre}{"dyn_d"}{$k}*$var{$k};
    }
    if ($k eq "constante")
    {
      $prop=$prop+$coef{$genre}{"dyn_d"}{$k};
    }
  }
  foreach $k (keys %{$coef{$genre}{"dyn_m"}})
  {

```

```

    if ($k ne "constante")
    {
        $logmontant=$logmontant+$coef{$genre}{"dyn_m"}{$k}$var{$k};
    }
    if ($k eq "constante")
    {
        $logmontant=$logmontant+$coef{$genre}{"dyn_m"}{$k};
    }
}

$gamma[$t]=$prop;
$beta[$t]=$logmontant;

#Tirage des chocs temporels hétéroscédastiques

$eps=Norm2;
$eta=exp($coef{$genre}{"dyn_disp"}{"lambda0"}
        +$coef{$genre}{"dyn_disp"}{"lambda_bin"}*$var{"lag_bin"}
        +$coef{$genre}{"dyn_disp"}{"lambda_log"}*$var{"lag_log"})*
($coef{$genre}{"dyn_disp"}{"rho_2"}*$eps+
sqrt(1-$coef{$genre}{"dyn_disp"}{"rho_2"}**2)*Norm2);

#Imputation
$parti[$t]=($prop+$ef1+$eps>0?1:0);
$logmontanti[$t]=$parti[$t]*($logmontant+$ef2+$eta);
}
}

#La fonction renvoie 4 tableaux :
#2 pour les imputations (recours et log-montant)
#2 pour les seuls predcteurs
#Elle renvoie également deux scalaires :
#Les deux effets fixes
return(@parti,@logmontanti,@gamma,@beta,$ef1,$ef2);
}

```

## 6.2 Analyse des trajectoires

## 7 Conclusion

## References

- [1] Albouy V, Davezies L., Debrand T. (2009): Dynamic estimation of health expenditure : A new approach for simulating individual expenditure, *Document de travail de l'IRDES*, n° 20.
- [2] Amemiya T. (1985): *Advanced Econometrics*, Basil Blackwell, Oxford
- [3] Bago d'Uva T. (2005): Latent class models for utilisation of health care, *Health Economics*, 14(9), pp. 873-92.
- [4] Bolhaar J., Lindeboom M. and van der Klaaw B. (2008): A dynamic analysis of demand for health insurance and Health care, *IZA Discussion paper*, DP 3698.
- [5] Buntin MB. and Zaslavsky AM. (2004): Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures, *Journal of Health Economics*, 23, pp. 525-42.
- [6] Cameron A. and Trivedi P. (1998): *Regression analysis of count data*, Econometric Society Monographs, Cambridge University Press.
- [7] Contoyannis P. , Jones A. and Rice N. (2004): The dynamics of health in the British household panel survey, *Journal of Applied Econometrics*, 19(4), pp. 473-503.
- [8] Dormont B., Grignon M. and Huber H. (2006): Health expenditure growth: reassessing the threat of ageing, *Health Economics*, 15, pp. 947-63.
- [9] Dormont B. (2009): Les dépenses de santé : une augmentation salubre ?, *Collection du CEPREMAP*.
- [10] Duan N. (1983): Smearing estimate: a nonparametric retransformation method, *Journal of the American Statistical Association*, 78, pp. 605-10.
- [11] Duan, N., Manning, W.G., Moris C. and Newhouse JP., (1983): A comparison of alternative models for the demand for medical care, *Journal of Business and Economics Statistics*, 1, pp. 115-26.
- [12] Grossman, M. (1972): On the concept of health capital and the demand for health, *Journal of Political Economy*, 80, pp. 223-55
- [13] Hay JW., Leu R. and Rohrer P. (1987): Ordinary least squares and sample-selection models of health-care demand, *Journal of Business Economics and Statistics*, 5, pp. 499-506.
- [14] Hay JW. and Olsen RJ. (1984): Let them eat cake: a note on comparing alternative models of the demand for medical care, *Journal of Business Economics and Statistics*, 2, pp. 279-82.
- [15] Heckman J. (1979): Sample selection bias as a specification error, *Econometrica*, 47, pp. 153-61.
- [16] Heckman J. (1981): The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process, in Manski and McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

- [17] Heckman J., Vytlacil E. (2005): Structural equations, Treatment effects, and econometrics policy evaluation, *Econometrica*, 73, pp. 669–738.
- [18] Honoré B., Kyriazidou E. (2000): Panel Discrete Choice Models with Lagged Dependent Variables, *Econometrica*, 68, pp. 839–874.
- [19] Jones A. (2000): Health econometrics, In: Culyer and Newhouse (Eds.), *Handbook of Health Economics*. Elsevier.
- [20] Jones AM., Rice N., Contoyannis P. (2006): The dynamics of health, In: Jones A. (eds), *Elgar Companion to Health Economics*, pp. 17-25.
- [21] Kyriazidou E. (1997): Estimation of a panel data sample selection model, *Econometrica*, 65 pp. 1335-1364.
- [22] Leung SF. and Yu S. (1996): On the choice between sample selection and two-part models, *Journal of Econometrics*, 72, pp. 197–229.
- [23] Maddala GS. (1985): *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, 3.
- [24] Magnac T. (2000): Subsidised Training and Youth Employment : Distinguishing Unobserved Heterogeneity from State Dependence in Labour Market Histories, *The Economic Journal*, 110, pp. 805–837.
- [25] Manning WG. and Mullahy J. (2001): Estimating log models: to transform or not to transform?, *Journal of Health Economics*, 20, pp. 461–94.
- [26] Manning WG. (1998): The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, 17, pp. 283–95.
- [27] Manning WG., Duan N. and Rogers WH. (1987): Monte Carlo evidence on the choice between sample selection and two-part models, *Journal of Econometrics*, 35, pp. 59–82.
- [28] Manski CF. (2003): Partial Identification of Probability Distributions, *Springer Series in Statistics*.
- [29] Marical F. (2007): En quoi la prise en compte des transferts liés à la santé modifie-t-elle l'appréciation du niveau de vie ?, *INSEE-France Portrait Social-Edition 2007*.
- [30] Mullahy, J. (1986): Specification and testing of some modified count data models, *Journal of Econometrics*, 33, pp. 341-65.
- [31] Mullahy J. (1998): Much ado about two: reconsidering retransformation and the two-part model in health econometrics, *Journal of Health Economics*, 17, pp. 247–81.
- [32] Nolan A. (2007): A dynamic analysis of GP visiting in Ireland: 1995-2001, *Health Economics*, 16, pp. 129-43.
- [33] Self SG., Liang KY. (1987): Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions, *Journal of the American Statistical Association*, 82, pp. 605–610.

- [34] Seshamani M. and Gray AM. (2004): A longitudinal study of the effects of age and time to death on hospital costs, *Journal of Health Economics*, 23, pp. 217–35.
- [35] Seshamani M. and Gray AM. (2004): Aging and health-care expenditure: the red herring argument revisited. *Health Economics*, 13, pp. 303-14.
- [36] Stearns S. and Norton E. (2004): Time to include time to death? the future of health care expenditure predictions, *Health Economics*, 13, 315–27.
- [37] Verbeek M. and Nijman T. (1992): Testing for selectivity bias in panel data models, *International Economic Review*, 33(3), pp. 681–704.
- [38] Wooldridge J. (2002): *Econometric Analysis of Cross-Section and Panel Data*, MIT Press: Cambridge.
- [39] Wooldridge J. (2005): Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics*, 20(1), pp. 39–54.
- [40] Zweifel P., Felder S. and Meiers M. (1999): Ageing of the population and health care expenditure: a red herring? *Health Economics*, 8, pp. 485–96.
- [41] Zweifel P., Felder S. and Werblow A. (2004): Population aging and health care expenditures: new evidence on the red herring, *The Geneva Papers on Risk and Insurance*, 29(4), pp. 652-66.

Table 5: Dépense ambulatoire pour les hommes

	TPM en coupe		Dépendance d'état	
	$\gamma$	$\beta$	$\gamma$	$\beta$
constante	1,45**	5,72**	1,37**	5,17**
recours en t-1 ( $D_{it-1}$ )			-0,37**	-0,94**
log-montant en t-1 $\ln(M_{it-1})$			0,25**	0,25**
condition initiale, recours( $\widehat{r}_{D_{i0}}$ )			0,22**	0,26**
condition initiale, log-montant( $\widehat{r}_{\ln(M_{i0})}$ )			0,09**	0,31**
age	-0,04**	-0,03**	-0,06**	-0,02
age <sup>2</sup> /100	0,20**	0,08*	0,16	0,02
age <sup>3</sup> /1000	-0,34**	0,05	-0,17	0,13
age <sup>4</sup> /10000	0,24**	-0,08*	0,09	-0,12**
ttd1=Mort l'année suivante	2,29**	1,10**	2,94*	0,69*
ttd2=Mort dans deux ans	1,66*	1,19**	1,88	0,88*
ttd1×age	-0,04**	-0,01*	-0,04*	-0,01
ttd2×age	-0,03**	-0,02**	-0,03*	-0,01*
independant	-0,68**	-0,43**	-0,44**	-0,33**
cadre du privé	0,07	0,10**	0,02	0,10**
fonctionnaire	-0,18**	0,02	-0,10	0,01
chômeur	-0,23**	0,09**	-0,19**	0,08
retraité	-0,15	-0,02	-0,09	-0,06
autre inactifs	0,12	0,65**	0,04	0,53**
jeune	-0,50**	-0,16	-0,77**	-0,21*
pers. de ref. cadre du privé (jeune)	-0,01	0,23**	-0,05	0,16**
pers. de ref. service public (jeune)	-0,11	0,17**	-0,11	0,12
pers. de ref. indépendant (jeune)	-0,26**	-0,02	-0,34**	0,03
pers. de ref. chômeur (jeune)	-0,35**	-0,17**	-0,29**	-0,17
pers. de ref. retraité (jeune)	-0,31**	-0,42**	-0,09	-0,37*
pers. de ref. autre inactif (jeune)	-0,15	-0,17*	-0,16	-0,14
primaire (adulte : atteint)	-0,04	-0,03	0,04	-0,03
premier cycle (adulte : atteint)	0,04	-0,07**	0,07	-0,03
supérieur (adulte : atteint)	0,10*	-0,05	0,10	-0,02
autre scolarité (adulte : atteint)	0,39**	0,03	0,22	0,05
non scolarisé (adulte : atteint)	-0,15	-0,36**	-0,08	-0,31**
primaire (jeune : actuel)	0,78**	0,15	0,73**	0,23*
collège (jeune : actuel)	0,49**	0,18*	0,43**	0,19
supérieur (jeune : actuel)	-0,36**	0,01	0,33**	0,18
autre scolarité (jeune : actuel)	3,72	0,95*	2,01	0,22
non scolarisé (jeune : actuel)	0,85**	0,20*	0,59**	0,22
$\rho_1$			0,64**	
$\rho_2$			0,19**	
$\sigma_v$			0,54**	
$\sigma_u$			0,30**	
$\sigma_\eta$	1,23**			
$\lambda_0$			0,21**	
$\lambda_{D_{t-1}}$			0,51**	
$\lambda_{\ln(M_{t-1})}$			-0,14**	

Table 6: Dépense ambulatoire pour les femmes

	TPM en coupe		Dépendance d'état	
	$\gamma$	$\beta$	$\gamma$	$\beta$
constante	2,56**	5,58**	2,23**	4,80**
recours en t-1 ( $D_{it-1}$ )			-0,67**	-0,71**
log-montant en t-1 $\ln(M_{it-1})$			0,39**	0,28**
condition initiale, recours( $\widehat{r}_{D_{i0}}$ )			0,12**	0,14**
condition initiale, log-montant( $\widehat{r}_{\ln(M_{i0})}$ )			0,01	0,30**
age	-0,06**	0,06*	-0,13**	0,03**
age <sup>2</sup> /100	-0,17**	-0,20**	0,31**	-0,10**
age <sup>3</sup> /1000	-0,20	0,38**	-0,33	0,20**
age <sup>4</sup> /10000	0,11	-0,21**	0,14	-0,11**
ttd1=Mort l'année suivante	1,96	1,40**	0,18	1,85**
ttd2=Mort dans deux ans	-0,80	0,54	-0,33	0,58
ttd1×age	-0,03	-0,02**	0,00	-0,02**
ttd2×age	0,01	-0,01	0,00	0,00
independant	-0,92**	-0,25**	-0,45**	-0,17
cadre du privé	0,19**	0,11**	0,09	0,05
fonctionnaire	-0,04**	-0,04	0,03	-0,03
chômeur	-0,24**	0,03	-0,11	0,02
retraité	0,10	-0,05	0,14	-0,01
autre inactifs	-0,37**	-0,11**	-0,38**	-0,11**
jeune	-1,17**	-0,86**	-1,12**	-0,80**
pers. de ref. cadre du privé (jeune)	0,13*	0,12**	0,09	0,07
pers. de ref. service public (jeune)	-0,01	0,01	-0,02	-0,03
pers. de ref. indépendant (jeune)	-0,29**	0,06	-0,25**	0,04
pers. de ref. chômeur (jeune)	-0,21**	-0,14**	-0,07	-0,11
pers. de ref. retraité (jeune)	-0,33**	-0,48**	0,00	-0,55**
pers. de ref. autre inactif (jeune)	0,16	-0,26**	0,44**	-0,22**
primaire (adulte : atteint)	-0,03	-0,01	0,07	0,04
premier cycle (adulte : atteint)	-0,02	-0,04	0,03	-0,01
supérieur (adulte : atteint)	-0,13**	-0,05	-0,12	-0,03
autre scolarité (adulte : atteint)	-0,39**	-0,15**	-0,21	-0,09
non scolarisé (adulte : atteint)	-0,08	-0,11*	0,09	-0,03
primaire (jeune : actuel)	0,56**	0,38**	0,27**	0,50**
collège (jeune : actuel)	0,40**	0,59**	0,19*	0,60**
supérieur (jeune : actuel)	0,01	0,52**	0,44**	0,67**
autre scolarité (jeune : actuel)	1,01**	0,88**	0,78*	0,85**
non scolarisé (jeune : actuel)	0,29*	0,66**	-0,26	0,57**
$\rho_1$			0,14**	
$\rho_2$			0,14**	
$\sigma_v$			0,42**	
$\sigma_u$			0,04**	
$\sigma_\eta$	1,11**			
$\lambda_0$			0,25**	
$\lambda_{D_{t-1}}$			0,46**	
$\lambda_{\ln(M_{t-1})}$			-0,14**	