IRDES

Document de travail
Working paper

Dynamic Estimation of Health Expenditure: A new approach for simulating individual expenditure

Valérie Albouy (INSEE)
Laurent Davezies (INSEE)
Thierry Debrand (IRDES)

DT n° 20

Janvier 2009

Dynamic Estimation of Health Expenditure: A new approach for simulating individual expenditure

Valerie Albouy*, Laurent Davezies† and Thierry Debrand‡

January 27, 2009

Abstract

This study compares estimates of outpatient expenditure computed with different models. Our aim is to predict annual health expenditures. We use a French panel dataset over a six year period (2000-2006) for 7112 individuals. Our article is based on the estimations of five different models. The first model is a simple two part model estimated in cross section. The other models (models 2 to 5) are estimated with selection models (or generalized tobit models). Model 2 is a basic sample selection model in cross section. Model 3 is similar to model 2, but takes into account the panel dimension. It includes constant unobserved heterogeneity to deal with state dependency. Model 4 is a dynamic sample selection model (with lagged adjustement), while in model 5, we take into account the possible heteroskedasticity of residuals in the dynamic model.

We find that all the models have the same properties in the cross section dimension (distribution, probability of health care use by gender and age, health expenditure by gender and age) but model 5 gives better results reflecting the temporal correlation with health expenditure. Indeed, the retransformation of predicted log transformed expenditures in homoscedastic models (models 1 to 4) generates very poor temporal correlation for " heavy consumers ", although the data show the contrary. Incorporation of heteroskedasticity gives better results in terms of temporal correlation.

Keywords: Health econometrics, expenditures, panel data, selection models.
Code JEL: I0, C1, C5.

The authors thank Francesco Moscone and Alberto Holly for their support and Badi Baltagi, William Greene, Jean-Paul Chaze, Andrew Jones and the participants of the Health Econometrics Workshop organised by the University of Leicester and the University of Milan-Bicocca for their helpful comments. We also thank Nelly Le Guen, Stephanie Guillaume and the entire ESPS team at IRDES. Usual disclaimers applied.

*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 Introduction

The economics of health and medical care has long been conceptualized as a dynamic process (Jones, 2000; Jones, Rice and Contoyannis, 2006). As early as 1972, Grossman described health as being a dimension of human capital in which people can invest. Individuals are endowed with an initial stock of health capital that depreciates through time. In Grossman's model, investment in health and the rate of health stock depreciation are not constant through lifecycles. The depreciation rate depends on the very nature of the agents themselves and on their social and economic environment. Indeed, where people live, how they live, what they eat and whether they are socially integrated or not, determine the rate of depreciation of their health capital. The depreciation rate also increases with age. As a consequence, to keep their health capital above a critical threshold, individuals have to increase their investment and thus their spending on health care. This results in an optimal individual (intertemporal) path of healthcare use that depends on income, level of education and many socioeconomic characteristics. There are also medical reasons for analyzing health expenditure in a dynamic process. Indeed, illnesses may persist for a long time or have long term consequences. Good understanding of health status and individual healthcare expenditure patterns requires taking account of individual past histories. With reference to Grossman's model, investment in health depends on the stock of capital built up during previous periods, thereby leading to the study of states of dependence.

One major problem with state of dependence is determining to what extent it explains the temporal correlation of health expenditure. Health status and health care use through time differ greatly among individuals. This heterogeneity can be explained by considerable unobserved heterogeneity that exceeds simple inaccuracy when measuring illnesses and health problems. Moreover, illnesses do not have the same consequences for all individuals; neither do treatments have the same effects, so that both health status and health expenditure vary substantially from one individual to another for a given illness. These differences in health consumption patterns may be explained by objective factors (gender, social origin, education level, income, family structure), while others are far more difficult to understand (for example, the genetic propensity of each individual to stand pain). This unobserved and time invariant individual heterogeneity leads to a "spurious dependence" that is difficult to disentangle from state dependence. The economic literature provides well-established evidence that state dependence and unobserved heterogeneity have different implications regarding public policy. As for health policies, it necessary to either develop preventive care (heavy dependency and low individual heterogeneity) for the entire population, or target health and social policies on the "weakest" persons (heavy individual heterogeneity).

In addition to these implications in terms of public policy recommendations, the relative importance of state dependence and individual heterogeneity is in fact vital from a descriptive standpoint. The purpose of this study is to assess non structural equations that provide individual predictors of healthcare consumption paths. Decomposing the mechanisms of temporal correlation is therefore crucial. The quality of simulations will heavily depend on the respective roles devoted to individual heterogeneity and to state of dependency.

This paper is divided into three parts. We first give a brief overview of the debates that have stirred health economists and econometricians who are concerned with cross-section expenditure estimates. We then explain why working with a panel changes the terms of the debate and propose a new dynamic approach to simulate individual health expenditure. Lastly, we estimate ambulatory expenditure on the basis of a panel of French data (ESPS) for the period 2000-2005.

2 Cross section estimation: Extensive debate

Debates on cross section modelling of health expenditure focus on two main issues. The first issue is relative to the fact that many people do not have any health expenditure during a given period of time. Health economists have wondered whether or not it is necessary to model the decision-making process leading to the amount of use observed as a joint decision-making process (i.e. decision to use on the one hand; how much to spend on the other). The second issue stems from the highly skewed distribution of health expenditure. Economists have often used logarithmic transformations which raises the problem of retransforming to the original scale.

2.1 Sample selection model or Two part model

Distributions of health expenditures contain a high proportion of zero observations (in France, over a one year period, only 7% have no use of ambulatory care). Health economists have adopted several strategies for dealing with this zero expenditure.

The first strategy is to ignore the accumulation point and to reason as if dealing with a continuous variable by performing linear regression, usually on quantity $log(M+1)$. This has the advantage of simplicity, but the problem is that it does not permit modelling the decision to seek care or not. Indeed, the probability of having no health expenditure - $P(M=0|X)$ - is equal to 0 in such a model. Furthermore, according to Duan (1983), the estimators obtained using this method have quite poor properties.

Another strategy consists in considering that we are dealing with a variable censored at zero. In this model, the continuous variable M depends on covariables X^M :

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

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

This model is a type I-Tobit model in the classification proposed by Amemiya in 1985. In such a model, the determinants of seeking health care and the determinants of positive health expenditure are the same. It also assumes that the coefficients linked to these determinants are the same. This is a very strong assumption.

The third method of dealing with the problem of zeros is to situate oneself in a classical model with truncated data and write a joint decision process (or sample selection model SSM or type II-tobit model in Amemiya's classification). In this case two distinct equations are written: the first determines whether the individual will have positive health care use, the second determines its amount. We denote as D_{it} the fact that individual i has to use on date t ($D_{it}=1$ in the case of use and $D_{it}=0$ if not). We note the amount of expenditure as M_{it}. In the case where there is no ambiguity, indices i and t are omitted. We note as X^D all the covariables related to D and X^M those related to M. The set of regressors X^D and X^M is noted X. We do not necessarily assume that $X^D \neq X^M$ although this case will be presented in what follows. The expense is deduced from the data (D_{it}, M_{it}):

$$D = \mathbf{1}_{\{X^D\gamma+\varepsilon>0\}} \Rightarrow P(D=1|X) = F_{-\varepsilon}(X^D\gamma) \quad (3)$$
$$M = (X^M\beta+\eta)D \quad (4)$$

In order to take into account the correlation between the residuals ε in the participation decision equation and the residuals η in the expenditure equation, the following assumption is made classically:

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

The fourth way of modeling health expenditure in health economics literature is known as the Two Part Model (2PM). It is modeled as a function of the decision to demand treatment. The assumption (5) of SSM is replaced by:

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

This model is very different from a selection model since in the Two Part Model the second equation only models $M|X, D=1$ and not the entirety of a latent distribution $M|X$ that would only be observed in the case $D=1$. This means that in the 2PM, we assume that $E(\eta|D=1,X)=0$, contrary to the SSM in which we assume that $E(\eta|D=1,X) = E(\eta|X\gamma+\varepsilon>0,X) = \lambda(X) \neq 0$. The coefficients β of SSM and 2PM cannot be compared since these two models do not assess the same underlying economic model. In the 2PM, the coefficients β cannot be interpreted as causal impacts of X on amount M (since unobserved heterogeneity influences participation D). Despite this difference of interpretation (which mainly affects the economic interpretation of β), it is possible to identify, estimate and simulate the conditional distribution[1] of $(D,M)|X$ using a 2PM as:

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

The choice between 2PM and SSM has provoked heated debate between the researchers of RAND (Duan et al., 1983; Manning, Duan and Rogers, 1987), who are partisans of 2PM, and other researchers (Hay, Leu and Rohrer, 1987; Hay and Olsen, 1984; Maddala, 1985) who support SSM type modeling. In brief, the pro-2PM criticise the fact that estimation in a SSM model relies on distributional assumptions (assumption 5) whose validity is often not tested (Duan et al., 1983). Furthermore, estimation greatly depends on this distributional assumption in the case where the covariates for D and M are the same; the robustness of SSM estimators is often considered poor if lacking exclusion variables[2]. The supporters of SSM question the existence of a bivariate distribution of (ε,η) which would be such that variable $(\eta|\varepsilon > -X^D\gamma)$ is normally distributed (Hay et Olsen, 1984). They emphasize that if the residuals of equations (3) and (4) are not independent (and they are probably not), the distribution of $(\eta|\varepsilon > -X^D\gamma)$ is a function of $X^D\gamma$. There is then no reason for $E(\eta|\varepsilon > -X^D\gamma) = 0$. Duan et al. (1983) find a counterexample to the first

[1] We note as d, m, and l the potential realization of the random variables D, M and $ln(M)$. We use d_{it}, m_{it} and l_{it} in the case where it is necessary to specify the individual and the time period considered. We misuse language to a certain extent by assimilating conditional distribution $\mathcal{L}_{D,M|X}(d,m)$ with likelihood $\mathcal{L}_{D,M|X}(D,M)$.

[2] The term exclusion variable is used when speaking about a variable Z such as $Z \in X^D$ and $Z \notin X^M$

argument above. They find a distribution allowing a correlation between ε and η, but which is such that η on the sub-sample of individuals demanding treatment are normally distributed and are independent of covariates common to the two equations.

Maddala (1985) entered the debate in favor of SSM, considering the model better represented the underlying decision process. But the partisans of 2PM argued that a sequential decision process reflected reality better. Maddala stressed that if the variables are jointly omitted in the two equations (such as health status), the residuals are necessarily correlated. Finally, he qualified the counter-example provided by Duan as "semantic". Duan et al. (1985) answered these arguments by underlining the fact that they were not interested in the coefficients β per se but in the prediction of average expenditure, i.e. prediction of $E(M|X)$ from:

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

Their simulations show that when the purpose is to predict $E(M|X)$ as accurately as possible, 2PM outperforms SSM which can be subject to multi-collinearity problems (Manning, Duan and Rogers, 1987; Leung and Yu, 1996). This result did not win consensus among economists but is consistent with the fact that the estimation of $E(M|X, D = 1)$ does not rely on the parametric specification of the joint distribution of (ε, η) in a 2PM, whereas this parametric specification is required for an SSM in the absence of a substantial support instrument [3](Manski, 1977; 1985; 1986; 2003).

An SSM model should be used in order to estimate the effect of variable X, all other things being equal (including health status, which is not observed) on the amount of health expenditure. However, it appears that many health economists are more interested in predicting health expenditure than in estimating a structural model. In such a case, the covariation of X and health expenditure cannot be interpreted as a causal effect, only as a descriptive one. But this does not matter if the purpose is to predict expenditures. This is the argument put forward by Manning, Duan and Rogers in 1987.

2.2 The transformation and retransformation problem

The second central issue in the literature devoted to econometric methods applied to health economics is the "retransformation problem" [4] (Manning, 1998, Mullahy, 1998; Manning and Mullahy, 2001). In equation (4), variable M can be the amount of health care expenses or any injective transformations of them. Many economists use a logarithmic transformation of expenses. In a linear model, working on health expenditure or logged expenditure does not require the same underlying assumptions. When working directly on the amounts, economists assume that covariates X^M and residuals η combine additionally, whereas when working on the logarithm of the amounts, they assume that the covariates

[3] This term is used to describe variable Z so that $Z \in X^D$ and $Z \notin X^M$ and so that $P(D = 1|Z, X^D)$ describe the whole of segment $[0, 1]$ when Z varies on its support. A large support instrument is therefore a specific exclusion variable.

[4] For the sake of simplicity, the models considered for continuous variables will be linear. A priori it is possible to consider non linear models though in this case the number of covariates must be kept low in order to avoid problems related to the inflation in the number of observations required to obtain precise estimations. The aim is no longer to estimate a vector in $\mathbb{R}^k$ but a function belonging to a much "bigger" functional space (the curse of dimensionality).

and residuals combine multiplicatively. More generally, they can work with any Box-Cox transformation of M (Chaze, 2005), $M \rightarrow \frac{(M-1)^{\lambda}}{\lambda}$ for $\lambda \geq 0$. The logged variable corresponds to $\lambda = 0$ while working directly on the variable corresponds to $\lambda = 1$; case $\lambda = \frac{1}{2}$ corresponds to a "quadratic" combination of the factors; case $\lambda = \frac{1}{3}$ corresponds to a "cubic" combination, etc. As the distribution of health expenditure is thick tailed, we chose a model in which the combination of factors is multiplicative rather than additive (λ is close to 0 and not 1). In what follows, we therefore work on log transformed health expenditure[5].

Logarithmic transformation raises a problem as the model estimates the conditional expectation of $E(ln(M)|D = 1, X^M)$ (or more generally the conditional law $L_{ln(M)|D=1,X^M}$), when the quantity of interest is $E(M|D = 1, X^M)$ (respectively $L_{M|D=1,X^M}$). With an additional technical assumption on the distribution of $\lambda|X^M, D = 1$ [6] , we obtain:

$$ln(M) = (X^M\beta + \eta)D \Rightarrow E(M|X^M, D = 1) = e^{X^M\beta} \int e^x dF_{\eta|X^M,D=1}(x)$$

Thus:

$$E(M|X^M, D = 1) = e^{X^M\beta} \left(e^{E(\eta|X^M,D=1)} + \sum_{k=1}^{+\infty} \frac{E\left[\left(\eta - E(\eta|X^M, D = 1)\right)^k |X^M, D = 1\right]}{k!} \right) \quad (6)$$

It should be recalled that in the case of 2PM, we have the following mean independence hypothesis: $E(\eta|X^M, D = 1) = 0$. If in addition we assume that $\eta|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$, then $\int e^x dF_{\eta|X^M,D=1}(x)$ is equal to $e^{\frac{1}{2}\sigma^2_\eta}$. More generally, with the mean independence assumption $E(\eta|X^M, D = 1) = 0$, the conditional normality assumption $\eta|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$ and even the conditional independence assumption $\eta \perp\!\!\!\perp X^M|D = 1$ are not required to estimate β. On the other hand, to estimate $E(M|X^M, D = 1)$, it is necessary to estimate $\int e^x dF_{\eta|X^M,D=1}(x)$ which depends on the distribution of η conditionally to X^M and $D = 1$ $F_{\eta|X^M,D=1}$ and not only on conditional expectation $E\left(\eta|X^M, D = 1\right)$. In particular, if we wrongly assume that $\eta|X^M, D = 1 \sim \mathcal{N}(0, \sigma^2)$, $e^{\frac{1}{2}\sigma^2_\eta}$ is a biased estimator of $\int e^x dF_{\eta|X^M,D=1}(x)$. Thus Duan (1993) proposed using the following empirical counterpart to estimate $\int e^x dF_{\eta|X^M,D=1}(x)$:

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

The estimator $\frac{1}{N}\sum_{i=1}^{N} e^{\eta_i}$ converges to $\int e^x dF_{\eta|X^M,D=1}(x) = E(e^\eta) = E(e^\eta|X)$ when $\eta \perp\!\!\!\perp X^M|D = 1$, therefore the conditional normality is no longer necessary. However, as equation (6) shows, if the centered moments of $\eta|X^M, D = 1$ depend on X^M, then the estimator is biased (since in this case $E(e^\eta) \neq E(e^\eta|X)$). Duan's estimator is therefore biased by heteroskedasticity ($V\left(\eta|X^M, D = 1\right)$ depends on X^M). For this reason, some researchers have recommended against working in log.

In this retransformation problem, Mullahy (1998) proposes estimating the second equation of 2PM with a generalized linear model, i.e. posing as hypothesis: $E\left(M|D = 1, X^M\right) = \exp\left\{X^M\beta\right\}$.

[5]However, what we state in the following remains valid for any other non linear transformation

[6]For example: $\eta \perp\!\!\!\perp X^M|D = 1$ and the sequence $\frac{E[(\eta - E(\eta|D=1))^n|D=1]}{n!}$ is bounded.

In their article "Estimating log models: to transform or not to transform?" Manning and Mullahy propose a series of tests to decide which model to choose. In brief, in the case of a thick tailed distribution, they recommend working on the logarithm of expenditure; otherwise they recommend opting for a generalized linear model to avoid the problem of retransformation.

Another alternative consists in explicitly modeling the moments of η, for example by assuming that: $\eta | X^M, D = 1 \sim \mathcal{N}(0, e^{X^M \lambda})$. This approach has the advantages and drawbacks of parametric methods: it is simple, makes simulations easy, but requires strong parametric assumptions.

We are now going to see that taking into account individual heterogeneity and dependency on longitudinal data requires using an SSM rather than a TPM, as well as using parametric methods. We have therefore decided to deal with heteroskedasticity by parametric modeling. As a consequence, the work presented in the rest of this article is "done in log".

3 Estimation on longitudinal data

3.1 People are not all alike

Most consumption behavior is characterized by considerable individual heterogeneity. Working on panel data makes it possible to take this heterogeneity into account more efficiently to describe, identify, estimate and simulate trajectories. Indeed, stable differences of behavior between individuals lead to thinking that non observed characteristics of individuals are constant through time.

Let us assume that conditionally upon observables X, we obtain the following model:

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

Quantities u_i and v_i are interpreted as follows: some individuals with the same characteristics X have to use treatment more frequently than others, and among individuals who have to use treatment, some consume more than others (with the same characteristics X). However, this difference between individuals cannot be observed directly; it can only be estimated on the basis of frequency of health care use and the amounts observed per individual. This permits considering correlations between events occurring at different dates: we omit assumption $(D_{it}, M_{it}) \perp\!\!\!\perp (D_{it'}, M_{it'}) | X$ which is unrealistic. It is especially important to omit this assumption if the final goal is to simulate trajectories in order to perform evaluations on life cycles.

We assume that couples (u_i, v_i) are independent and identically distributed. What is more, we assume that they are independent from X , thus we use a "random effect" panel data model. Theoretically, this is not essential: we could just as well use a "fixed effect" model without an assumption on joint distribution (u_i, v_i, X). Nevertheless this would raise estimation difficulties linked to the problem of incident parameters. Kyriazidou (1997) proposed a method of estimating asymptotically without biasing parameters β et γ by using observations such as $\gamma_{it} X_{it}^D = \gamma_{it'} X_{it'}^D$. However, this method is not necessarily legitimate when considering the substantial effects of age. In addition, heterogeneity u_i, v_i is treated by Kyriazidou as a nuisance parameter but the estimation of its distribution is not considered, whereas it is a key element in a descriptive approach. Our aim is to perform simulations on the life cycle, thus simulate quantities u_i and v_i. If we estimate a fixed effects model, it is also necessary to estimate the correlations between u_i, v_i and X to make a consistent imputation. By using a model with random effect, this projection is performed directly, but coefficients γ and β cannot be interpreted as the causal impact of X if certain non observed characteristics constant through time are simultaneously correlated to X and $(D, ln(M))$. Once again this is not necessarily problematic when taking a descriptive approach to perform the micro-simulation on the life cycle. Lastly, to our knowledge there have been no empirical applications of Kyriazidou's method, thus information on its empirical difficulties is relatively scare. Consequently, we reserve the implementation of this strategy for future works.

As in a cross section model we assume that the couples $(\varepsilon_{it}, \eta_{it})$ are independent and identically distributed and independent of X. Lastly, we assume that $(u_i, v_i) \perp\!\!\!\perp (\varepsilon_{it}, \eta_{it})|X$.

3.2 Individual heterogeneity points to privileging SSM

It remains to be determined whether we formulate assumptions of 2PM type $E(v_i + \eta_{it}|X_{it}^M, D_{it} = 1) = 0$ or SSM type (dependence of variables $u_i, v_i, \varepsilon_{it}, \eta_{it}$). As with $E(v_i + \eta_{it}|X_{it}^M, D_{it} = 1) = E(v_i|X_{it}^M, u_i > -X_{it}^D\gamma - \varepsilon_{it}) + E(\eta_{it}|X_{it}^M, \varepsilon_{it} > -X_{it}^D\gamma - u_i)$, the aim is to specify the assumptions we intend making on the joint laws of u and v on the one hand and on that of ε and η on the other. Discussion relating to the pertinence of assumption $\varepsilon_{it} \perp\!\!\!\perp \eta_{it}|X$ versus $\varepsilon_{it}, \eta_{it}|X \sim \mathcal{N}\left((0,0), \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right)$ has been reported in 2.1.

Assumption $u_i \perp\!\!\!\perp v_i|X$ implies that the amount consumed on a date is independent (conditionally on X) of the number of visits during all the periods observed. This result is fairly intuitive since the non observed variables u_i imply a greater or lesser frequency of use, whereas the non observed variables v_i imply greater or lesser consumption. The formal proof of the result can be found in the appendix 7.1. This assumption is clearly rejected by the data, as consumption by period of individuals who often use health care is much higher than that of other individuals. Thus there is a correlation between u_i and v_i (given X), naturally leading us to prefer an SSM model rather than a 2PM model in the framework of longitudinal estimation.

In the case of selection models and in the absence of a large support instrument, it is usual (and necessary) to make parametric hypotheses concerning the distribution of non observed heterogeneity. To estimate the model we assume [7] thus now that:

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

With this hypothesis, it is possible to estimate the model either by maximum likelihood, or by two step procedures (estimation of parameters γ by random effect probit then estimation of β by integration of a Mills ratio). The two-step estimation method cannot be used with a state of dependency. Since we intend to introduce such a dependency in our model (cf. section 3.3), we privilege estimation by maximum likelihood.

[7] As usual, the variance of ε can be normalized to 1 since parameter γ can only be identified up to scale.

When reasoning over a period and for an individual, the likelihood of observation $D_{it}, ln(M_{it})$ is therefore $f_{ln(M_{it})|X_{it},u_i,v_i,D_{it}=1}(ln(M_{it}))P(D_{it}=1|X_{it},u_i,v_i)D_{it} + P(D_{it}=0|X_{it},u_i,v_i)(1-D_{it})$. The likelihood of an individual can then be deduced easily:

$$\mathcal{L}_{(D,ln(M))_{t=1...T}|X}(d_t,l_t)_{t=1...T} =$$
$$\int\int\left[\prod_{t=1}^{T} f_{ln(M_t)|X,u,v,D_t=1}(l_t)P(D_t=1|X,u,v)d_t + P(D_t=0|X,u,v)(1-d_t)\right]\phi(u,v)dudv$$

However in our case:

$$\begin{aligned} f_{ln(M_t)|X,u,v,D_t=1}(l_t) &= f_{\eta_t|X,u,v,\varepsilon_t>-X_t\gamma-u}(l_t - X_t\beta - v) \\ &= \frac{e^{\frac{-(l_t-X_t\beta-v)^2}{2\sigma_\eta^2}} * \Phi\left(\frac{1}{\sqrt{1-\rho_2^2}}\left(X_t\gamma + u + \rho_2\frac{(l_t-X_t\beta-v)}{\sigma_\eta}\right)\right)}{\sqrt{2\pi}\sigma_\eta\Phi\left(X_t\gamma+u\right)} \end{aligned}$$

We then obtain:

$$\mathcal{L}_{(D,ln(M))_{t=1...T}|X}(d_t,l_t)_{t=1...T} = \int\int\left[\prod_{t=1}^{T} f_{it}(u,v)\right]\phi(u,v)dudv$$

With:

$$\begin{aligned} f_{it}(u,v) &= \frac{1}{\sqrt{2\pi}\sigma_\eta}e^{\frac{-(l_t-X\beta-v)^2}{2\sigma_\eta^2}} * \Phi\left(\frac{1}{\sqrt{1-\rho_2^2}}\left(X_t\gamma + u + \rho_2\frac{(l_t - X_t\beta - v)}{\sigma_\eta}\right)\right)d_t \\ &\quad + \left(1 - \Phi(X_t\gamma + u)\right)(1-d_t) \end{aligned}$$

And :

$$\phi(u,v) = \frac{1}{2\pi\sqrt{1-\rho_1^2}\sigma_u\sigma_v}e^{\frac{\sigma_u^2v^2+\sigma_v^2u^2-2\rho_1\sigma_u\sigma_vuv}{2\left(1-\rho_1^2\right)\sigma_u^2\sigma_v^2}}$$

.

3.3 State dependence and the problem of initial conditions

Grossman's theoretical health model supposes state dependence for healthcare expenditure. Although many authors work on panel datasets, to our knowledge very few papers deal with estimations of the dynamic health expenditure model (Jones and Rice, 2004; Bago d'Uva, 2005; Nolan, 2007).

Formally, the model considered is written in the following way[8] for the date $t > 0$:

$$D_{it} = \mathbf{1}_{\{D_{it-1}\gamma_1 + ln(M_{it-1})\gamma_2 + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}} \quad (7)$$
$$ln(M_{it}) = (D_{it-1}\beta_1 + ln(M_{it-1})\beta_2 + X_{it}\beta + v_i + \eta_{it})D_{it} \quad (8)$$

If we want to obtain a structural interpretation of the model, we are confronted with a major difficulty when estimating coefficients $\gamma_1, \gamma_2, \gamma, \beta_1, \beta_2, \beta$ since, by nature, the couple $D_{it-1}, ln(M_{it-1})$ is correlated to couple u_i, v_i. This constitutes a barrier if we want to obtain a "good" description of paths D_{it}, M_{it} in order to simulate it. Reasoning in the same way as for covariates X does not permit overcoming the problem. Indeed, to perform

[8] We use the convention $log(0) = 0$. Thus if $D_{it-1} = 0$ then $log(M_{it-1}) = 0$

a "good" simulation, (u_i, v_i) and D_{it-1}, M_{it-1} have to be jointly simulated. By applying this argument to previous dates, the problem that emerges is that of jointly simulating individual heterogeneity (u_i, v_i) and initial conditions $D_{i0}, ln(M_{i0})$.

Indeed, the distribution of paths is broken down as follows:

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

However, equations (7) et (8) imply that $(D_0, ln(M_0))$ is correlated to (u, v). Therefore it is not possible to write the joint law $\mathcal{L}_{D_0,ln(M_0),u,v|X}$ as a product of a marginal distribution:

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

Estimating such a dynamic model is therefore difficult insofar as the joint distribution of $D_0, ln(M_0), u, v$ is not easily identifiable. Furthermore, putting forward the assumption that $D_0, ln(M_0)$ is independent of (u, v) is in contradiction with the model. This problem reflects the difficulty of separating state dependence and individual heterogeneity. The initial conditions are endogenous.

3.3.1 Heckman's approach (1981)

Heckman (1981) studied the problem of endogeneity of initial conditions in a binary model with state dependence. If we extrapolate this method to the case dealt with here, the aim is to approach distribution $D_0, ln(M_0)|u, v, X$ by postulating that the linked latent variables in the first period approximate a relation of type:

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

This relation is not structural since parameters $\gamma_0, \beta_0, \mu_i (i = 1...4)$ have no economic meaning. In addition, even the functional form used for modeling the law of $D_0, ln(M_0)|u, v, X$ is not consistent with the dynamics (as opposed to the linear model).

If we admit that the likelihood of $D_0, ln(M_0)|u, v, X$ really takes the form assumed previously, the likelihood of the model is written as:

$$\mathcal{L}_{(D,ln(M))_{t=0...T}|X}(d_t, l_t)_{t=0...T} = \int\int \left[\prod_{t=1}^{T} f_{it}(u, v)\right] \cdot f_{i0}(u, v) \cdot \phi(u, v) dudv$$

With:

$$\begin{aligned} f_{it}(u,v) \quad = \quad & \frac{1}{\sqrt{2\pi}\sigma_\eta} e^{\frac{-(l_{it}-d_{it-1}\beta_1-l_{it-1}\beta_2-X_{it}\beta-v)^2}{2\sigma_\eta^2}} . \\ & \Phi\left(\frac{1}{\sqrt{1-\rho_2^2}}\left(\begin{array}{c} d_{it-1}\gamma_1 + l_{it-1}\gamma_2 + X_{it}\gamma + u \\ +\rho_2\frac{(l_{it}-d_{it-1}\beta_1-l_{it-1}\beta_2-X_{it}\beta-v)}{\sigma_\eta} \end{array}\right)\right) d_{it} \\ & + (1 - \Phi(d_{it-1}\gamma_1 + l_{it-1}\gamma_2 + X_{it}\gamma + u)) (1 - d_{it}) \end{aligned}$$

And:

$$f_{i0}(u,v) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma_0} e^{\frac{-(l_{i0}-X_{i0}\beta_0-\mu_3 u-\mu_4 v)^2}{2\sigma_0^2}} .$$

$$\Phi\left(\frac{1}{\sqrt{1-\rho_0^2}}\left(X_{i0}\gamma_0+\mu_1 u+\mu_2 v+\rho_0\frac{(l_{i0}-X_{i0}\beta_0-\mu_3 u-\mu_4 v)}{\sigma_0}\right)\right)d_{i0}$$
$$+\left(1-\Phi(X_{i0}\gamma+\mu_1 u+\mu_2 v)\right)(1-d_{i0})$$

In practice this likelihood is quite difficult to estimate: it is difficult to "separate" the β, the γ, the β_0 and the γ_0. What is more, the convergence of the estimators is ensured only when $T \rightarrow +\infty$. In his works on the binary model, Heckman highlighted that the estimators on the simulated data can be strongly biased in the case of few periods. As seen in what follows, we have a panel of 6 periods that makes using such a method inappropriate.

3.3.2 Wooldridge's approach (2005)

Wooldridge (2005) took a different approach from Heckman. Instead of making an assumption on the conditional distribution of $D_0, ln(M_0)|u, v, X_0$, he proposed making an assumption on the distribution $(u, v)|D_0, ln(M_0), (X_t)_{t=0...T}$.

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

This approach is very close to that of Chamberlain for getting round the problem of incident parameters in the case of non linear panel data models. The heterogeneity is broken down between a correlated part with X (term of type X_iK_X) and a non correlated part modeled by a normal random effect. In this case the aim is not to control a possible correlation between the heterogeneity and X but between the heterogeneity and the initial conditions given X (term of type $D_{i0}K_D + ln(M_{i0})K_M + X_iK_X$).

By observing that we now have:

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

It is possible to estimate the following model for $t > 0$:

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

and estimate the distribution of $D_0, ln(M_0)|X$ in cross-section to simulate the initialisation of the trajectory (term $\mathcal{L}_{D_0,ln(M_0)|X}$).

One of the main advantages of Wooldridge's method lies in its ease of implementation. Variables $D_0, ln(M_0)$ and $(X_t)_{t=0...T}$ therefore play exactly the same role in the estimations as covariables X, and the estimation techniques presented previously apply (maximum likelihood, two-step estimation and corrections using Mills ratio, etc.). The non observed heterogeneity terms observed through time are: $D_{i0}\kappa_{uD}+ln(M_{i0})\kappa_{uM}+\sum_{t=0}^{T}X_{it}\kappa_{uX_{it}}+u_i$ and $D_{i0}\kappa_{vD}+ln(M_{i0})\kappa_{vM}+\sum_{t=0}^{T}X_{it}\kappa_{vX_{it}}+v_i$. Components $D_{i0}\kappa_{uD}+ln(M_{i0})\kappa_{uM}+\sum_{t=0}^{T}X_{it}\kappa_{uX_{it}}$ and $D_{i0}\kappa_{vD}+ln(M_{i0})\kappa_{vM}+\sum_{t=0}^{T}X_{it}\kappa_{vX_{it}}$ is correlated to the initial conditions and terms u_i and v_i represent the heterogeneity components that are not correlated with the other variables. Another advantage, contrary to the procedure proposed by Heckman, is that the convergence of the estimators for $T < +\infty$ is assured (provided that the parametric hypothesis on heterogeneity assumption u, v is verified).

However, the number of parameters to be estimated can increase considerably (the size of K_X is $K \times T$), moreover quasi collinearity problems can occur between X and $(X_t)_{t=0...T}$ and between $D_{t-1}, ln(M_{t-1})$ and $D_0, ln(M_0)$. Regarding our data, it was impossible to separate the different effects convincingly by using this method, mainly due to the correlation of age and consumption on date $t = 0$ with age and consumption on the following dates. Therefore we propose an adaptation of Wooldridge's method in the following section to get round these difficulties.

3.3.3 An adaptation of Wooldridge's method

Let us recall the approach used in Wooldridge's method. It essentially entails assuming that the heterogeneity takes the form:

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

Where f belongs to a set of functions. In his original version, Wooldridge considered the set of linear functions f. This set is therefore quite large if there are a lot of variables X and a lot of periods. The impossibility of separating the effects of manner convincingly in our data can be explained by excessive flexibility in modeling the non observed heterogeneity.

Consequently, we propose reformulating Wooldridge's assumption by considering a more limited set of functions f than the set of linear functions. $u, v|D_0, ln(M_0), X$ must represent the over-propensity (and under-propensity respectively) of using health care and the over-propensity (and under-propensity respectively) of having high expenditures taking into account the individual's characteristics X. The problem of initial conditions is precisely that this over or under propensity to expend is correlated with the over or under propensity to consume on date $t = 0$. We therefore propose using the generalized residual of r_0 which is the best available estimation of the over or under propensity to expend on the initial date. In a linear framework, the generalized residue would simply be the "classical" residual $Y_0 - E[Y_0|X_0]$. In the framework of a logit or probit model $(Y_0 = \mathbf{1}_{\{X_0\delta+\xi\}})$ the generalized residual would be $E\left(\xi|Y_0, X_0\right)$.

This leads us to define the data generating process for a cross-section model. We used a 2PM in $t = 0$ to estimate the generalized residuals for the following reasons:

- A data generating process of SSM type in $t = 0$ is not compatible with the dynamic model chosen for $t > 0$
- The absence of exclusion variable makes the estimations of an SSM on the sole first period very unstable
- This estimation in $t = 0$ is not structural. We do not want to interpret parameters γ_0 and β_0 as causal impacts, therefore utilization of an SSM is not essential. The main aim is to take a descriptive and robust approach for estimating the over or under propensity to expend $r_{D_0}, r_{ln(M_0)}$ on date $t = 0$ of an individual with characteristics X_0. From a descriptive standpoint, the 2PM is more robust to heteroskedasticity and to the non normality of the residuals in the amounts equation.

The generalized residual is written as[9] thus:
For a consumer in t=0:

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

For a non consumer in t=0:

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

It is therefore possible to postulate that individual heterogeneity is distributed according to the following conditional law:

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

3.4 The problem of heteroskedasticity

As the problem of the initial conditions has been solved, it is now necessary to take into account the possible heteroskedasticity of residuals in a dynamic model. The descriptive statistics highlight strong heteroskedasticity in the amounts logarithm: indeed the cloud of points representing the log of the amounts of two successive consumptions is considerably flattened for the large amounts (see figures 1 and 2).

The omission of this heteroskedasticity is liable to bias the estimation of coefficients β and γ (contrary to the linear model), and it also raises retransformation problems, since the average level of the amounts is linked to the dispersion of log-amounts (as $E(e^{\eta}) = e^{\frac{\sigma^2}{2}}$ in the Gaussian case). It is therefore also necessary to model possible heteroskedasticity. This is what we have done by assuming that:

$$\eta|D_{t-1}, m_{t-1}, X_t \sim \mathcal{N}(0, e^{2(\lambda_0 + \lambda_D D_{t-1} + \lambda_{ln(M_{t-1})} ln(M_{i,t-1}))})$$

The model with heteroskedasticity becomes:

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

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

[9] In our problem, if the data generating process on date $t=0$ is an SSM, the generalized residuals are:
For a consumer in t=0:

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

For a non consumer in t=0:

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

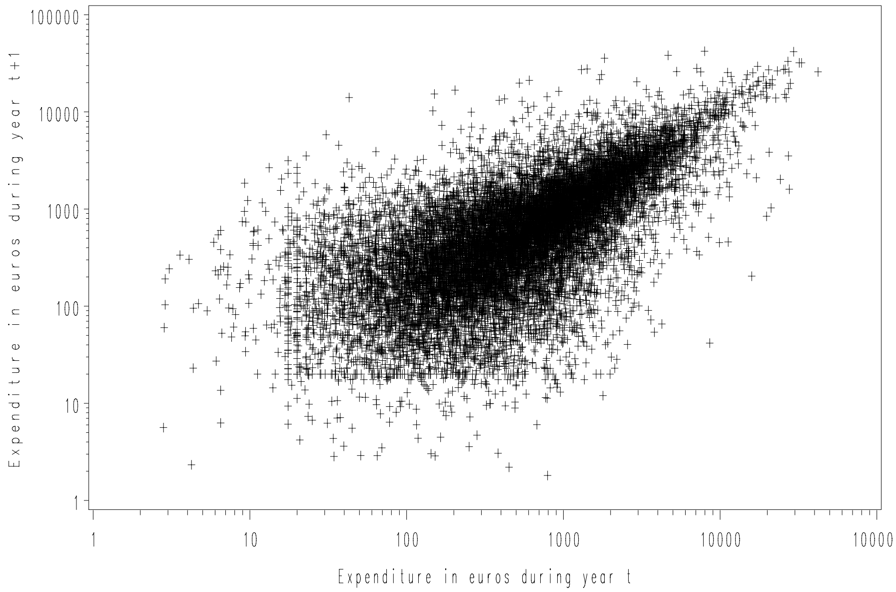

Figure 1: Expenditure in t and $t+1$ for men (logarithmic scale)

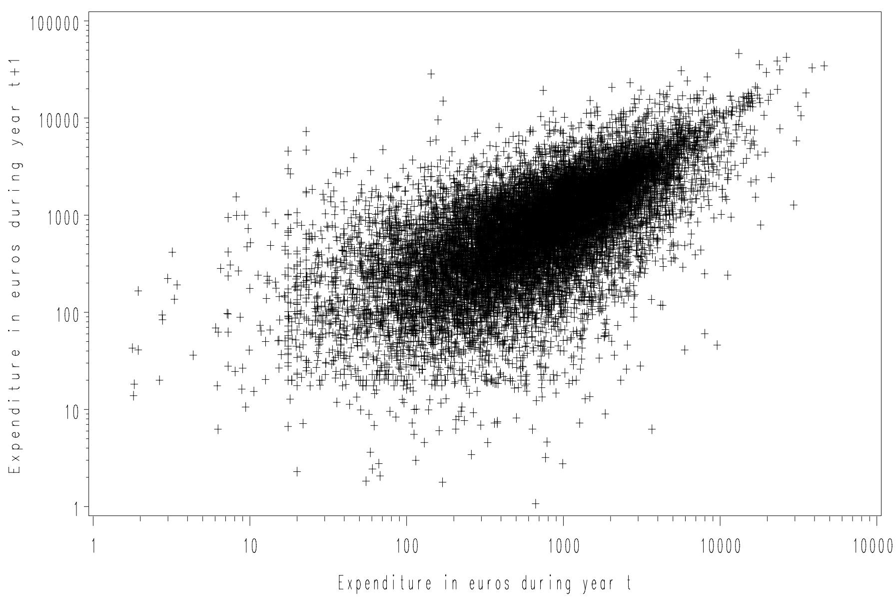

Figure 2: Expenditure in t and $t+1$ for women (logarithmic scale)

The covariance variance matrix of $(u_i, v_i, \varepsilon_{it}, \eta_{it})$ is :

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

and $\widehat{r}_{ln(M_{i0})} = \widehat{E}(\tilde{\varepsilon}_{i0}|D_{i0}, ln(M_{i0}), X_{i0})$ and $\widehat{r}_{D_{i0}} = \widehat{E}(\tilde{\eta}_{i0}|D_{i0}, ln(M_{i0}), X_{i0})$ are the estimators of the generalized residuals of the following 2PM:

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

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

4 Data and descriptive statistics

4.1 The specificities of ESPS – EPAS

The empirical work uses matched administrative and survey data. The survey data on which this study is based are from the IRDES[10] Social Welfare and Health Surveys (ESPS). ESPS provides detailed information relative to the socio-economic situation and the health status (disability, self reported health, health status index). We matched these data with health expenditure data from public national insurance: the Permanent Health Insurance Samples (EPAS). For every person in the ESPS 2000 dataset, we collected health expenditures (amount and utilization) from EPAS for the period 2000-2006. This provides a panel dataset over 2000-2006 with 7 112 individuals. This panel is unbalanced due to several reasons. In very particular cases, individuals may have changed health insurance system (civil servants have a different insurance system), or have died. We used the ESPS 2004 survey to determine whether people were dead (as it has a retrospective influence on health care use).

Our panel data is quite unusual in the sense that socioeconomic data are only available for the year 2000 while only health expenditures are available for the following six years. Consequently, all the characteristics in our model are time invariant (except age and time to death), which is problematic when disentangling them from constant unobserved heterogeneity. More precisely, Socioeconomic variables, such as gender and level of education, are invariant through time while others vary mechanically, such as age, number of years before decease (if deceased). The only variables we were able to modify were occupational status and marital status. We assume that these remained constant through the period, thus we take them as invariants stemming from the ESPS 2000 survey. As for the health expenditure variables, they vary and change year by year, since they stem from the EPAS files of 2000-2001-2002-2003-2004-2005.

4.2 Descriptive statistics

Table 1 gives simple statistics for the health expenditure distribution in our data. The third column gives the proportion of individuals who used health care during the year. The fourth column gives the mean health expenditure amounts: 93% of the individuals of our panel had an average annual consumption of €1.107. Concerning socio economic characteristics, for occupational status, 44% of individuals worked whereas 15% were retired, 25% were students and 16% were "inactive". The latter category groups both the unemployed and persons not seeking employment (excluding students and pensioners). The probability of employed individuals who used ambulatory care is comparable to that of inactive individuals (93%), although the amount of this ambulatory expenditure is lower (€952 vs €1 375).

[10]IRDES: Institute for research and information on health economics

Table 1: Panel data

	2000 ESPS-EPAS	2001-2003 EPAS	2004 ESPS-EPAS	2005 EPAS
Socio-économic Characteristics	Age Gender Employment status Education level Household structure		reasons for Attrition	
Health expenditures	Ambulatory inpatient	Ambulatory inpatient	Ambulatory inpatient	Ambulatory inpatient

Table 2: Statistic descriptives

	Ind.	ambulatory		Nb of death
		$D = 1$	$\overline{m}$ in €	
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
All	41 831	0,93	1107	345

The economic literature of the last two decades clearly highlights the importance of time to death as one of the key variables for understanding the healthcare costs of individuals. In our panel, we are able to observe health expenditure one and two years before the decease of individuals. Indeed, the variability of health care use is considerable between individuals who will die in the next year (who spend €4201 in ambulatory care) and those who will live more than two years afterward (who spend €1 067).

These descriptive statistics show great variability in health expenditure levels: the first quartile is €171 and the third is €1 308. The left tail is longer. 55.1% men and 41.2% women spend more than €500 a year whereas only 3.1% of men and women spend more than €5 000 per year. 0.3% of the individuals spend more than €15 000 a year in outpatient care. If we select only individuals alive at the end of 2005 and examine their consumption year after year for the entire period, 76.3% of these individuals have ambulatory expenses every year and almost 9 out of 10 individuals have expenses at least 5 times in the 6 periods observed. Very few persons had no ambulatory consumption for the entire period (0.1%). The amount of expenses is correlated to their frequency: the median of the amount of consumptions is €817 for those who had consumed every year versus less than €200 for those who had consumed less than one year in two.

4.2.1 The variables selected for our estimations

We selected the following covariates for our estimations:

- AGE: a polynomial of order 4 for age.
- TIME TO DEATH: dummy variables, one and two years before death. This allows taking into account the acceleration of health expenditure before death. These two variables are crossed with age.
- OCCUPATIONAL ACTIVITY: this variable only concerns adults who are independent professionals, private sector managers, civil servants, unemployed, pensioners, other inactive persons. The reference modality is private sector employee (non managerial).
- LEVEL OF EDUCATION: this corresponds to the highest level of education obtained: primary, Bachelor degree, higher, other education. The reference is a Master's degree.
- YOUNG: this dummy is used to take into account the fact that a child in a household still lives with his or her parents. Thus for this child the activity of the head of the household is taken into account according to the modalities mentioned previously. As for level of studies the current level is taken into account according to the modalities mentioned above.

5 Results

5.1 The models estimated:

In the following we discuss how the five models fit the data. The first model is a basic "Two part model" estimated in cross section, as presented in 2.1. This model takes no advantage of the panel dimension. It enables accurate cross-section predictions but this

Table 3: Health expenditures

	Nb	ambulatory			
		$D = 1$	$\overline{m}$ (in €)		
Gender			Moy	Q1	Q3
Male	3 463	0,91	1 000	123	1 080
Female	3 649	0,94	1 207	241	1 491
Employment status					
Employee	3 122	0,93	952	178	1 151
Retiree	1 131	0,98	2 472	903	2 960
Out of employment	1 096	0,93	1 375	249	1 647
Student	1 763	0,89	422	81	544
Level of education					
other schooling	595	0,94	1 091	187	1 317
primary education	1 822	0,94	1 234	171	1 482
school	2 489	0,92	1 071	162	1 231
advanced	933	0,91	1 092	168	1 327
university	1 273	0,92	1 016	180	1 221
Age					
0-29	2 620	0,89	477	86	589
30-44	1 892	0,93	810	159	985
45-59	1 307	0,94	1 370	300	1 622
60-74	947	0,97	2 083	713	2 541
75 +	346	1,00	2 937	1 232	3 483
Time to death					
Alive		0,93	1 067	169	1 283
death in 2 years		0,94	3 417	636	4 405
death in 1 years		0,96	4 201	685	5 083
All	7 112	0,93	1 107	171	1 308

Table 4: Ambulatory expenditure by gender (€)

	Males		Females		All	
	%	$\overline{m}$	%	$\overline{m}$	%	$\overline{m}$
no expenditure	8,9	0	5,7	0	7,3	0
0 - 499 €	46,2	207	35,5	231	40,7	218
500 € - 999 €	18	724	20,8	730	19,5	728
1 000 € - 4 999 €	23,8	2 044	34,9	2 022	29,5	2 031
4 999 € - 14 999 €	2,7	7 743	2,8	7 460	2,7	7 594
15 000 € +	0,4	21 635	0,3	20 468	0,3	21 110
All	100	1 000	100	1 206	100	1 107

predictive performance is of no use for formulating public policy recommendations as it is not a structural model (Heckman and Vytlacil, 2005). However, as mentioned before, our aim is to predict medical care demand. Despite having made parametric hypotheses, this model can be identified semi-parametrically (the most straightforward 2PM is a parametric model but may be identified with semi parametric procedures) [11], which is an element of robustness. Nonetheless, the basic 2PM performs poorly when predicting health expenditure paths (see 3.2). Selection models (or type II tobit models according to Amemiya's classification) outperform the 2PM as they better predict correlation between utilization over life cycle and annual amount of health care received.

Models 2 to 5 are therefore selection models. All the specifications take into account unobserved heterogeneity that simultaneously influences the decision for utilization and the amount of expenditure. Model 2, i.e. the SSM presented in 2.1, does not use the panel dimension. This model is semi-parametric when using a large support instrumental variable (Manski, 2003). We found no such instrument. But the model can be identified under parametric assumptions (allowing characterization of functions $a \mapsto E(\eta|X, e > a)$). Some econometrists stress the lack of robustness of these models in the absence of an instrument (although this is controversial). This is particularly true when authors require structural equations. Without any instrument, it is difficult to disentangle causal effects of a covariate X in the decision of utilization and in the amounts of care received. However, this is less true when authors wish to predict and simulate medical care expenditure.

Model 3, presented in 3.2 makes use of the panel dimension. The specification contains time invariant unobserved heterogeneity (u and v). Residuals are assumed to be orthogonal to the covariates X. This assumption is too strong in a structural approach. Indeed, there is a risk of attributing part of individual heterogeneity to covariates X. A public policy rule that might modify the distribution of X would not however modify individual heterogeneity, which could bias evaluation. Conversely, this model is quite acceptable from our standpoint of lifecycle forecasting. Manski (1987) showed that although individual heterogeneity and temporal shocks are independent of covariates, parameter γ is semi-parametrically identifiable and estimable with a large support instrument. By assuming an exclusion restriction between the regressions, Kyriazidou (1997) proposed a semiparametric estimator for covariates that can remain constant from one period to another [12](no effect of age or period).

Model 4 is dynamic: we introduce lagged dependent variables. Introduction of lagged dependent variables with time invariant unobserved heterogeneity raises estimation problems: it is necessary to distinguish the relative influence of state dependence and unobserved heterogeneity. For linear models, Arellano and Bond proposed a semiparametric estimation method based on first differences and the use of lagged covariate values as instruments. For binary models, Honoré and Kyriazidou (2000) also proposed a semiparametric estimation method which requires no parametric assumption on the distributions of individual heterogeneity and of time varying residuals. However, this method is only valid for estimating effects of time-invariant regressors. Honoré and Kyriazidou (2000) and Magnac (2000) also studied the case of multinomial logit models, with no parametric assumption of the distribution of individual heterogeneity, but by specifying the distribution of temporal shocks. The method we use to deal with the problem of initial conditions

[11]By semi-parametrically we mean that the law of non observed u, v, ε and η is not specified parametrically. The models can only be identified by certain characteristics of the conditional distribution (for example: conditional moment, conditional median, etc.).

[12]The method proposed by Kyriazidou is based on hypotheses of type $P((X_{it}^D - X_{it'}^D)\gamma = 0) > 0$.

is due to Wooldridge (2005), though we adapt this method to improve the stability of the estimations (see 3.3.3).

Model 5 is model 4 but with the homoskedasticity assumption relaxed (in the second equation). The assumption of homoskedasticity is sensitive if the data have to be retransformed in the original scale (see 2.2). Furthermore, empirical evidence of heteroskedasticity is very strong (see 3.4). The highly parametric nature of model 4 facilitates its extension when heteroskedasticity depends on covariates or lagged endogenous variables.

	Model 1	Model 2	Model 3	Model 4	Model 5
Description	Two part model in cross-section	Sample selection model in cross-section	Sample selection model	Sample selection model	Sample selection model
Individual Heterogeneity	No	No	Yes	Yes	Yes
State dependence	No	No	No	Yes	Yes
Heteroskedasticity	No	No	No	No	Yes
Equation	$D_{it} = X_{it}\gamma + \varepsilon_{it}$ $M_{it} = X_{it}\beta + \eta_{it}$	$D_{it} = X_{it}\gamma + \varepsilon_{it}$ $M_{it} = X_{it}\beta + \eta_{it}$	$D_{it} = X_{it}\gamma + u_i + \varepsilon_{it}$ $M_{it} = X_{it}\beta + v_i + \eta_{it}$	$D_{it} = D_{it-1}\gamma_1 +$ $M_{it-1}\gamma_2 +$ $X_{it}\gamma + u_i + \varepsilon_{it}$ $M_{it} = D_{it-1}\beta_1 +$ $M_{it-1}\beta_2 +$ $X_{it}\beta + v_i + \eta_{it}$	$D_{it} = D_{it-1}\gamma_1 +$ $M_{it-1}\gamma_2 +$ $X_{it}\gamma + u_i + \varepsilon_{it}$ $M_{it} = D_{it-1}\beta_1 +$ $M_{it-1}\beta_2 +$ $X_{it}\beta + v_i + \eta_{it}$
Estimation assumption	$(\varepsilon,\eta)\|X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & 0\\0 & \sigma_\eta^2\end{bmatrix}\right)$	$(\varepsilon,\eta)\|X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho_2\sigma_\eta\\\rho_2\sigma_\eta & \sigma_\eta^2\end{bmatrix}\right)$	$(\varepsilon,\eta)\|X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho_2\sigma_\eta\\\rho_2\sigma_\eta & \sigma_\eta^2\end{bmatrix}\right)$ $(u,v)\|X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma_u^2 & \rho_1\sigma_u\sigma_v\\\rho_1\sigma_u\sigma_v & \sigma_v^2\end{bmatrix}\right)$ $(u,v) \perp\!\!\!\perp (\varepsilon,\eta)$	$(\varepsilon,\eta)\|D_{it-1}, M_{it-1}, X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho_2\sigma_\eta\\\rho_2\sigma_\eta & \sigma_\eta^2\end{bmatrix}\right)$ $(u,v)\|D_0, M_0, X \sim$ $\mathcal{N}\left(\begin{bmatrix}\overline{u}\\\overline{v}\end{bmatrix},\begin{bmatrix}\sigma_u^2 & \rho_1\sigma_u\sigma_v\\\rho_1\sigma_u\sigma_v & \sigma_v^2\end{bmatrix}\right)$ $\overline{u} = f(D_0, M_0, X)$ $\overline{v} = g(D_0, M_0, X)$ $(u,v) \perp\!\!\!\perp (\varepsilon,\eta)$	$(\varepsilon,\eta)\|D_{it-1}, M_{it-1}, X \sim$ $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1 & \rho_2\sigma_\eta\\\rho_2\sigma_\eta & \sigma_\eta^2\end{bmatrix}\right)$ $\sigma_\eta =$ $e^{\lambda_0+\lambda_D D_{it-1}+\lambda_M M_{it-1}}$ $(u,v)\|D_0, M_0, X \sim$ $\mathcal{N}\left(\begin{bmatrix}\overline{u}\\\overline{v}\end{bmatrix},\begin{bmatrix}\sigma_u^2 & \rho_1\sigma_u\sigma_v\\\rho_1\sigma_u\sigma_v & \sigma_v^2\end{bmatrix}\right)$ $\overline{u} = f(D_0, M_0, X)$ $\overline{v} = g(D_0, M_0, X)$ $(u,v) \perp\!\!\!\perp (\varepsilon,\eta)$
The identification assumption of γ if weaker than estimation assumption	$Med(\varepsilon\|X) = 0$ and Instrumental variable with large support	$Med(\varepsilon\|X) = 0$ and Instrumental variable with large support	$Med(\varepsilon\|X) = 0$ and Instrumental variable with large support		
The identification assumption of β if weaker than estimation assumption	$\eta \perp\!\!\!\perp \varepsilon$ $E(\eta\|X) = 0$	Instrumental variable with large support or the function $a \mapsto E(\eta\|X, \varepsilon > a)$ is known up to scale	Instrumental variable with large support or the function $a \mapsto E(\eta\|X, \varepsilon > a)$ is known up to scale		

5.2 Selection criteria

It is quite easy to select between these models. In fact these models are embedded: model 4 is a restriction of model 5 ($\lambda_{D_{t-1}} = \lambda_{ln(M_{t-1})} = 0$). Model 3 is a restriction of model 4 ($\gamma_1 = \gamma_2 = \beta_1 = \beta_2 = 0$). Model 2 is itself a restriction of model 3 ($\sigma_v = \sigma_u = 0$). Generally speaking, model 1 is not a restriction of model 2, but it becomes one if we assume that the hypothesis of the normality of η is maintained in model 1($\rho_1 = 0$). In order to test these nested models, we simply test the nullity of certain parameters. Classical tests supplied by software allow easy discrimination between the models. Nonetheless, caution is required when testing model 3 against model 2. The assumption we test is $\sigma_u = \sigma_v = 0$ versus $\sigma_u \geq 0$ and $\sigma_v \geq 0$, thus the test must take place at the bound (because a variance cannot be negative). The usual tests are not valid. A suitable test has been proposed by Self and Liang (1987), who establish that the classical test overestimates the p-value. Thus in our case, hypothesis $\sigma_v = 0$ or $\sigma_u = 0$ can be rejected by the classical test.

5.3 Estimation

The results of the estimations are reported in tables 7 and 8 in appendix 7.1. Health expenditure is correlated with age, "time to death" and social status as already stated by many authors(Zweifel, Felder and Meiers ,1999; Zweifel, Felder and Werblow, 2004; Stearns and Norton, 2004; Seshamani and Gray, 2004).

Models 2,3,4 and 5 show that the heterogeneity terms of the selection equation and of the amount received equation are correlated ($\rho_1 \neq 0$ and $\rho_2 \neq 0$). Models 3,4 and 5 show strong individual heterogeneity ($\sigma_u > 0$ and $\sigma_v > 0$),which could lead to spurious state dependence. However, models 4 and 5 show that non-spurious state dependence is significant in consumption dynamics.

Tests on the restrictions show that model 5 outperforms model 4 ($\lambda_{D_{it}} \neq 0$, $\lambda_{ln(M_{it})} \neq 0$). For similar reasons, model 4 is preferred to model 3 and model 3 is prefered to model 2 which is preferred to model 1.

Our aim is to simulate health expenditure paths. We check if they fit the paths observed in the ESPS-EPAS panel:

- For each model, the covariate is used to calculate predictor $X\widehat{\gamma}$ and $X\widehat{\beta}$.
- An amount of ambulatory expenditure is simulated for each individual. In order to do this, heterogeneity is sampled $(\varepsilon, \eta, u, v)$ in the estimated distribution.
- We then compare these simulations with the data.

The five distributions of simulated amounts have approximatively the same characteristics (density, probability of utilization by gender and age, amount by gender and age) in the cross-section dimension. But model 5 clearly outperforms the other models in the temporal dimension. Indeed, temporal correlations of healthcare amounts are stronger in model 5 (see tables 5,6), even if they remain underestimated compared to those observed in empirical data. Retransformation by the exponential function in homoskedastic models (model1-model4) generates very poor temporal correlations for "big consumers", though the data show this temporal correlation is strong. Introducing heterosckedasticity clearly improves the performance of the model to perform longitudinal simulations.

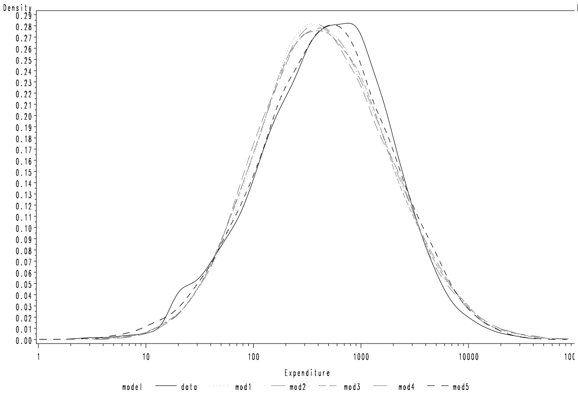

Figure 3: Density of amounts for men

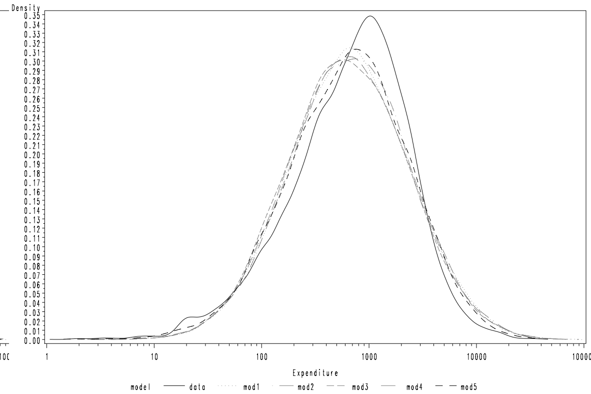

Figure 4: Density of amounts for women

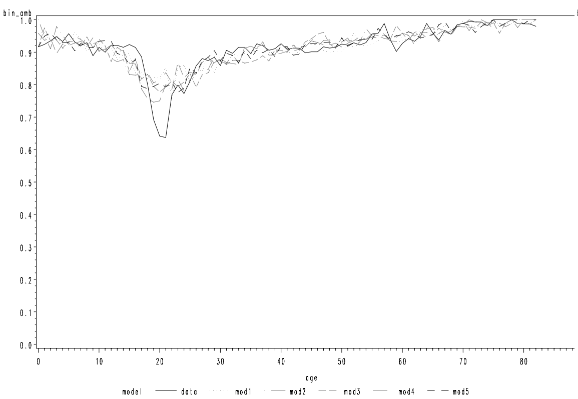

Figure 5: Probability of use by age for men

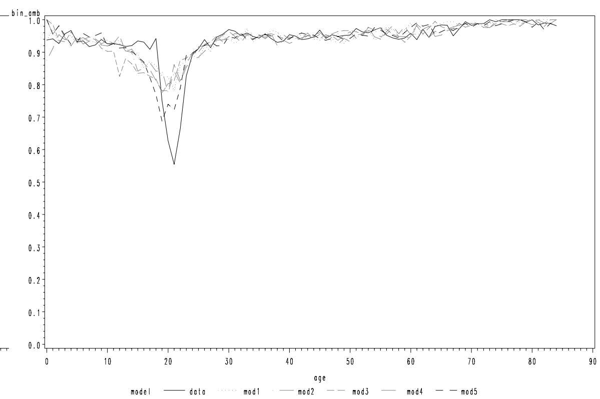

Figure 6: Probability of use by age for women

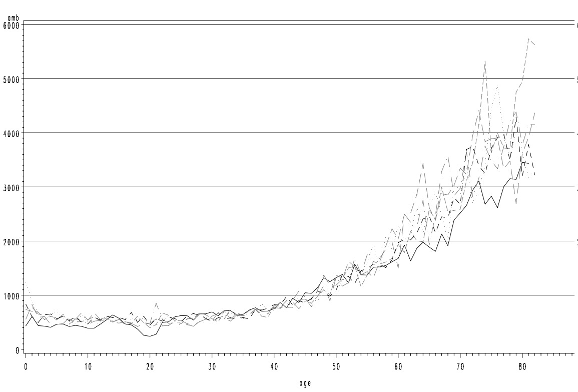

Figure 7: Average of expenditure by age for men

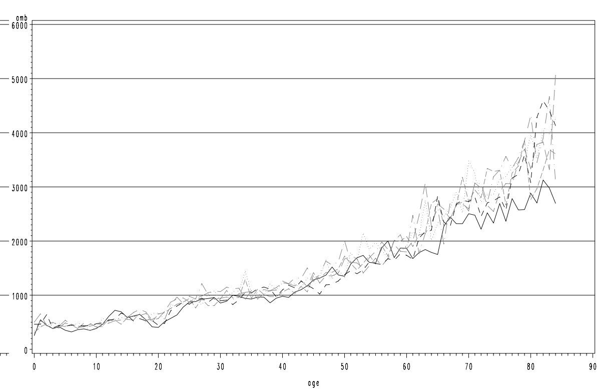

Figure 8: Average of expenditure by age for women

	Data	Model 1	Model 2	Model 3	Model 4	Model 5
$CORR(D(t), D(t-1))$	*0.34927*	0.05204	0.05956	0.28498	0.36384	0.36302
$CORR(D(t), M(t-1))$	*0.12686*	0.04037	0.04747	0.10053	0.10335	0.12591
$CORR(M(t), D(t-1))$	*0.11728*	0.04814	0.04585	0.09863	0.10596	0.11763
$CORR(M(t), M(t-1))$	*0.71493*	0.12213	0.10678	0.39960	0.46033	0.63117

Table 5: Correlation over the time of ambulatory expenditure, Men

	Data	Model 1	Model 2	Model 3	Model 4	Model 5
$CORR(D(t), D(t-1))$	*0.41418*	0.034138	0.03936	0.40561	0.43880	0.45571
$CORR(D(t), M(t-1))$	*0.13305*	0.040330	0.04702	0.10228	0.11219	0.13047
$CORR(M(t), D(t-1))$	*0.12622*	0.035377	0.03423	0.10241	0.11743	0.13005
$CORR(M(t), M(t-1))$	*0.68349*	0.089521	0.11953	0.39079	0.47964	0.58738

Table 6: Correlation over the time of ambulatory expenditure, Women

6 CONCLUSION

To conclude, the major difficulty in simulating healthcare consumption over a lifecycle is to reproduce the correlation of behavior over time. This paper highlights three implications:

- Two part models ignore the correlation (for each individual over time) between frequency of utilization and amount of expenditure.
- A good method of measuring constant individual heterogeneity and of state dependence is needed. What is more, the problem of endogeneity of initial condition must be treated.
- Logarithmic transformation could cause the simulation to be sensitive to the distribution of residuals, so great attention must be given to their heteroskedasticity.

In this paper we propose a strategy based on the estimation of a sample selection model using panel data with state dependence. Modeling heteroskedascticity improves the estimated correlation over time.

References

[1] Amemiya T. (1985): *Advanced Econometrics*, Basil Blackwell, Oxford

[2] Bago d'Uva T. (2005): Latent class models for utilisation of health care, *Health Economics*, 14(9), pp. 873-92.

[3] 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.

[4] 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.

[5] Cameron A. and Trivedi P. (1998): *Regression analysis of count data*, Econometric Society Monographs, Cambridge University Press.

[6] Chaze J-P. (2005) : Assessing household health expenditure with Box-Cox censoring models, 14, pp. 893-907.

[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] Duan N. (1983): Smearing estimate: a nonparametric retransformation method, *Journal of the American Statistical Association*, 78, pp. 605–10.

[10] 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.

[11] Grossman, M. (1972): On the concept of health capital and the demand for health, *Journal of Political Economy*, 80, pp. 223-55

[12] 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.

[13] 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.

[14] Heckman J. (1979): Sample selection bias as a specification error, *Econometrica*, 47, pp. 153–61.

[15] 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.), S*tructural Analysis of Discrete Data with Econometric Applications,* MIT Press.

[16] Heckman J., Vytlacil E. (2005): Structural equations, Treatment effects, and econometrics policy evaluation, *Econometrica*, 73, pp. 669–738.

[17] Honoré B., Kyriazidou E. (2000): Panel Discrete Choice Models with Lagged Dependante Variables, *Econometrica*, 68, pp. 839–874.

[18] Jones A. (2000): Health econometrics, In: Culyer and Newhouse (Eds.), *Handbook of Health Economics. Elsevier.*

[19] Jones AM., Rice N., Contoyannis P. (2006): The dynamics of health, In: Jones A. (eds), *Elgar Companion to Health Economics*, pp. 17-25.

[20] Kyriazidou E. (1997): Estimation of a panel data sample selection model, *Econometrica*, 65 pp. 1335-1364.

[21] Leung SF. and Yu S. (1996): On the choice between sample selection and two-part models, *Journal of Econometrics*, 72, pp. 197–229.

[22] Maddala GS. (1985): *Limited-Dependent and Qualitative Variables in Econometrics, E*conometric Society Monographs, 3.

[23] 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.

[24] Manning WG. and Mullahy J. (2001): Estimating log models: to transform or not to transform?, *Journal of Health Economics*, 20, pp. 461–94.

[25] Manning WG. (1998): The logged dependent variable, heteroskedasticity, and the retransformation problem. *Journal of Health Economics*, 17, pp. 283–95.

[26] 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.

[27] Manski CF. (2003): Partial Identification of Probability Distributions, *Springer Series in Statistics.*

[28] Mullahy, J. (1986): Specification and testing of some modified count data models, *Journal of Econometrics*, 33, pp. 341-65.

[29] 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.

[30] Nolan A. (2007): A dynamic analysis of GP visiting in Ireland: 1995-2001, *Health Economics*, 16, pp. 129-43.

[31] Self SG., Liang KY. (1987): Asymptotic Poperties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions, *Journal of the American Statistical Association*, 82, pp. 605–610.

[32] 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.

[33] Seshamani M. and Gray AM. (2004): Aging and health-care expenditure: the red herring argument revisited. *Health Economics*, 13, pp. 303-14.

[34] Stearns S. and Norton E. (2004): Time to include time to death? the future of health care expenditure predictions, *Health Economics*, 13, 315–27.

[35] Verbeek M. and Nijman T. (1992): Testing for selectivity bias in panel data models, *International Economic Review*, 33(3), pp. 681–704.

[36] Wooldridge J. (2002): *Econometric Analysis of Cross-Section and Panel Data*, MIT Press: Cambridge.

[37] 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.

[38] 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.

[39] 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.

7 Appendix :

7.1 Proof :

$$\mathcal{L}_{(D_t,ln(M_t))_{t=1\dots T}|X} = \mathcal{L}_{(D_t)_{t=1\dots T}|X} \cdot \mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T}}$$

$$\mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T}} = \int\int \mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T},u,v} \cdot \mathcal{L}_{u,v|X,(D_t)_{t=1\dots T}} dudv$$

Because (ε_t, η_t) are independent and identically distributed and independent of X, we have $\mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T},u,v} = \prod_{t=1}^{T} \mathcal{L}_{ln(M_t)|X,(D_t)_{t=1\dots T},u,v}$. Moreover, because $\eta_t \perp\!\!\!\perp \varepsilon_{t'}|X,u,v$ for $t \neq t'$, we have $\mathcal{L}_{ln(M_t)|X,(D_t)_{t=1\dots T},u,v} = \mathcal{L}_{ln(M_t)|X,D_t,u,v}$

$$\mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T}} = \int\int \prod_{t=1}^{T} \mathcal{L}_{ln(M_t)|X,D_t,u,v} \cdot \mathcal{L}_{u,v|X,(D_t)_{t=1\dots T}} dudv$$

We have $\mathcal{L}_{ln(M_t)|X,D_t,u,v} = \mathcal{L}_{ln(M_t)|X,D_t,v}$ because $\eta \perp\!\!\!\perp u|X,v$ and $\mathcal{L}_{v|X,(D_t)_{t=1\dots T},u} = \mathcal{L}_{v|X,u}$ because $v \perp\!\!\!\perp \varepsilon|X,u$.

$$\begin{aligned}\mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T}} &= \int\int \prod_{t=1}^{T} \mathcal{L}_{ln(M_t)|X,D_t,v} \cdot \mathcal{L}_{v|X,(D_t)_{t=1\dots T},u} \cdot \mathcal{L}_{u|X,(D_t)_{t=1\dots T}} dudv \\ &= \int\int \prod_{t=1}^{T} \mathcal{L}_{ln(M_t)|X,D_t,v} \cdot \mathcal{L}_{v|X,u} \cdot \mathcal{L}_{u|X,(D_t)_{t=1\dots T}} dudv\end{aligned}$$

If we assume $u \perp\!\!\!\perp v|X$, we have :

$$\mathcal{L}_{(ln(M_t))_{t=1\dots T}|X,(D_t)_{t=1\dots T}} = \int \prod_{t=1}^{T} \mathcal{L}_{ln(M_t)|X,D_t,v} \cdot \mathcal{L}_{v|X} dv \quad \textit{because} \quad \int \mathcal{L}_{u|X,(D_t)_{t=1\dots T}} du = 1$$

And thus by integration :

$$u \perp\!\!\!\perp v|X \Rightarrow \mathcal{L}_{ln(M_\tau)|X,(D_t)_{t=1\dots T}} = \mathcal{L}_{ln(M_\tau)|X,D_\tau}$$

So we can test assumption $u \perp\!\!\!\perp v|X$, by testing for example : $E\left(M_\tau|X, D_\tau = 1\right) = E\left(M_\tau|X, D_\tau = 1, \sum_{t=0}^{T} D_t\right)$

If $u \perp\!\!\!\perp v|X$, people who often use health care ($\sum_{t=0}^{T} D_t$ is high) must have the same level of amount of expenditure than other people with the same observable ccovariate X.

Table 7: Ambulatory expenditure for males

	model 1		model 2		model 3		model 4		model 5	
	γ	β	γ	β	γ	β	γ	β	γ	β
constant	1,45**	5,72**	1,56**	5,78**	2.42**	5.73**	0,73**	5,30**	1,37**	5,17**
utilization to ambulatory in t-1 (D_{it-1})							-0,31**	-0,89**	-0,37**	-0,94**
Amount of ambulatory expenditure in t-1 $ln(M_{it-1})$							0,25**	0,23**	0,25**	0,25**
initial condition, utilization ($\widehat{r}_{D_{i0}}$)							0,21**	0,24**	0,22**	0,26**
initial condition, amount($\widehat{r}_{ln(M_{i0})}$)							0,08**	0,30**	0,09**	0,31**
age	-0,04**	-0,03**	-0,05**	-0,04**	-0.06**	-0.02	0,00	-0,02*	-0,06**	-0,02
$age^2/100$	0,20**	0,08*	0,23**	0,10**	0.15	-0.02	-0,08	0,05	0,16	0,02
$age^3/1000$	-0,34**	0,05	-0,38**	0,02	-0.08	0.25**	0,19	0,06	-0,17	0,13
$age^4/10000$	0,24**	-0,08*	0,25**	-0,06	0.03	-0.19**	-0,10	-0,08	0,09	-0,12**
ttd1=time to death one year	2,29**	1,10**	1,20	1,34**	0.52	0.85**	2,60	0,62	2,94*	0,69*
ttd2=time to death two years	1,66*	1,19**	0,79	1,36**	0.27	0.84**	1,73	0,81*	1,88	0,88*
ttd1×age	-0,04**	-0,01*	-0,03*	-0,02	-0.01	-0.01	-0,04*	-0,01	-0,04*	-0,01
ttd2×age	-0,03**	-0,02**	-0,02	-0,02	-0.10	-0.01*	-0,03*	-0,01	-0,03*	-0,01*
self-employed	-0,68**	-0,43**	-0,69**	-0,43**	-0.92**	-0.54**	-0,44**	-0,33**	-0,44**	-0,33**
executive	0,07	0,10**	0,07	0,10**	0.07	0.11*	0,02	0,11**	0,02	0,10**
public service	-0,18**	0,02	-0,18**	0,02	-0.24**	-0.03	-0,11	0,00	-0,10	0,01
unemployed	-0,23**	0,09**	-0,23**	0,09**	-0.33**	0.05	-0,21**	0,06	-0,19**	0,08
retired	-0,15	-0,02	-0,13	-0,02	-0.40**	-0.27**	-0,13	-0,05	-0,09	-0,06
other non-working	0,12	0,65**	0,13	0,65**	0.26	0.61**	0,03	0,50**	0,04	0,53**
young	-0,50**	-0,16	-0,48**	-0,18**	-0.93**	-0.16	-0,73**	-0,23**	-0,77**	-0,21*
father executive (young)	-0,01	0,23**	-0,01	0,24**	0.01	0.23**	-0,05	0,16	-0,05	0,16**
father public service (young)	-0,11	0,17**	-0,11	0,17**	-0.03	0.10	-0,10	0,11	-0,11	0,12
father self-working (young)	-0,26**	-0,02	-0,26**	-0,01	-0.42**	-0.01	-0,35**	0,00	-0,34**	0,03
father unemployed (young)	-0,35**	-0,17**	-0,36**	-0,17**	-0.43**	-0.19	-0,31**	-0,16	-0,29**	-0,17
father retired (young)	-0,31**	-0,42**	-0,31**	-0,39**	-0.44	-0.52**	-0,09	-0,34*	-0,09	-0,37*
father other non-working (young)	-0,15	-0,17*	-0,15	-0,14	-0.24	-0.18	-0,16	-0,14	-0,16	-0,14
primary school (adult: attain)	-0,04	-0,03	-0,03	-0,03	-0.03	-0.08	0,02	-0,03	0,04	-0,03
school (adult: attain)	0,04	-0,07**	0,04	-0,07**	0.06	-0.07	0,08	-0,04	0,07	-0,03
advanced (adult: attain)	0,10*	-0,05	0,10*	-0,05	0.10	-0.04	0,11	-0,03	0,10	-0,02
other schooling career(adult: attain)	0,39**	0,03	0,41**	0,04	0.54*	0.04	0,20	0,05	0,22	0,05
no schooling (adult: attain)	-0,15	-0,36**	-0,18	-0,36**	-0.30	-0.46**	-0,12	-0,33**	-0,08	-0,31**
primary school (young: current)	0,78**	0,15	0,72**	0,15	1.05**	0.04	0,86	0,22*	0,73**	0,23*
school (young: current)	0,49**	0,18*	0,46**	0,19**	0.69**	0.09	0,42**	0,19*	0,43**	0,19
advanced (young: current)	-0,36**	0,01	-0,38**	0,03	-0.32	-0.04	0,28*	0,17	0,33**	0,18
other schooling career (young: current)	3,72	0,95*	0,92	1,06**	0.23	0.17	1,95	0,20	2,01	0,22
no schooling (young: current)	0,85**	0,20*	0,76**	0,18	1.03**	0.13	0,90**	0,18	0,59**	0,22
$\rho1$					0.67**		0,64**		0,64**	
$\rho2$					-0.02		0,05		0,19**	
σ_v			-0,01		0.86**		0,50**		0,54**	
σ_u					0.97**		0,28**		0,30**	
σ_η	1,23**		1,22**		0.90**		0,92**			
λ_0									0,21**	
$\lambda_{D_{t-1}}$									0,51**	
$\lambda_{ln(M_{t-1})}$									-0,14**	

Table 8: Ambulatory expenditure for females

	model 1		model 2		model 3		model 4		model 5	
	γ	β	γ	β	γ	β	γ	β	γ	β
constant	2,56**	5,58**	2,71**	5,49**	5.40**	5.75**	0,81**	5,08**	2,23**	4,80**
utilization to ambulatory in t-1 (D_{it-1})							-0,62**	-0,72**	-0,67**	-0,71**
Amount of ambulatory expenditure in t-1 $ln(M_{it-1})$							0,38**	0,26**	0,39**	0,28**
initial condition, utilization($\widehat{r}_{D_{i0}}$)							0,11**	0,15**	0,12**	0,14**
initial condition, amount($\widehat{r}_{ln(M_{i0})}$)							0,00	0,29**	0,01	0,30**
age	-0,06**	0,06*	-0,07**	0,06**	-0.16**	0.04**	0,00	0,02*	-0,13**	0,03**
age^2/100	0,17*	-0,20**	0,19*	-0,22**	0.34**	-0.13**	-0,09	-0,08*	0,31**	-0,10**
age^3/1000	-0,20	0,38**	-0,22	0,40**	-0.24	0.27**	0,20	0,18**	-0,33	0,20**
age^4/10000	0,11	-0,21**	0,11	-0,22**	0.06	-0.15**	-0,11	-0,11**	0,14	-0,11**
ttd1=time to death one year	1,96	1,40**	0,48	1,81**	0.33	1.74**	-0,19	1,71**	0,18	1,85**
ttd2=time to death two years	-0,80	0,54	-0,45	0,46	-0.38	0.08	-0,24	0,54	-0,33	0,58
ttd1×age	-0,03	-0,02**	-0,01	-0,02**	-0.01	-0.02	0,00	-0,02**	0,00	-0,02**
ttd2×age	0,01	-0,01	0,00	-0,01	0.00	0.00	0,00	0,00	0,00	0,00
self-employed	-0,92**	-0,25**	-0,93**	-0,24**	-1.49**	-0.32**	-0,44**	-0,17	-0,45**	-0,17
executive	0,19**	0,11**	0,18**	0,11**	0.26	0.11	0,09	0,06	0,09	0,05
public service	-0,04	-0,04	-0,05	-0,04	-0.04	-0.05	0,04	-0,03	0,03	-0,03
unemployed	-0,24**	0,03	-0,25**	0,04	-0.38**	0.02	-0,10	0,01	-0,11	0,02
retired	0,10	-0,05	0,09	-0,05	-0.08	-0.17**	0,06	-0,02	0,14	-0,01
other non-working	-0,37**	-0,11**	-0,38**	-0,11**	-0.61**	-0.16**	-0,38**	-0,12**	-0,38**	-0,11**
young	-1,17**	-0,86**	-1,21**	-0,83**	-2.49**	-0.86**	-1,01**	-0,80**	-1,12**	-0,80**
father executive (young)	0,13*	0,12**	0,13*	0,12**	0.20	0.13	0,07	0,07	0,09	0,07
father public service (young)	-0,01	0,01	-0,01	0,01	0.12	-0.03	-0,04	-0,04	-0,02	-0,03
father self working (young)	-0,29**	0,06	-0,29**	0,07	-0.34	0.03	-0,24**	0,03	-0,25**	0,04
father unemployed (young)	-0,21**	-0,14**	-0,21**	-0,14**	-0.25	0.17	-0,12	-0,12	-0,07	-0,11
father retired (young)	-0,33*	-0,48**	-0,34*	-0,49**	-0.65	-0.47*	0,02	-0,37*	0,00	-0,55**
father other non-working (young)	0,16	-0,26**	0,16	-0,26**	0.25	-0.21	0,43**	-0,26**	0,44**	-0,22**
primary school (adult: attain)	-0,03	-0,01	-0,03	-0,01	-0.09	-0.03	0,05	0,03	0,07	0,04
school (adult: attain)	-0,02	-0,04	-0,02	-0,04	-0.04	-0.04	0,01	-0,02	0,03	-0,01
advanced (adult: attain)	-0,13**	-0,05	-0,13**	-0,05	-0.25*	-0.05	-0,12	-0,02	-0,12	-0,03
other school career (adult: attain)	-0,39**	-0,15**	-0,39**	-0,15**	-0.56**	-0.20*	-0,21	-0,08	-0,21	-0,09
no schooling (adult: attain)	-0,08	-0,11*	-0,01	-0,13**	-0.22	-0.17*	-0,07	-0,02	0,09	-0,03
primary school (young: current)	0,56**	0,38**	0,52**	0,40**	0.82**	0.32**	0,65**	0,40**	0,27**	0,50**
school (young: current)	0,40**	0,59**	0,40**	0,59**	0.72**	0.58**	0,27*	0,56**	0,19*	0,60**
advanced (young: current)	0,01	0,52**	0,03	0,51**	0.24	0.59**	0,38*	0,77**	0,44**	0,67**
other school career (young: current)	1,01**	0,88**	1,04**	0,87**	1.06	0.67**	0,76*	0,74**	0,78*	0,85**
no schooling (young: current)	0,29*	0,66**	0,22	0,70**	0.09	0.51**	0,51**	0,43**	-0,26	0,57**
ρ_1					0.51**		0,07		0,14**	
ρ_2					-0.14**		0,05		0,14**	
σ_v			-0,02		0.78**		0,43**		0,42**	
σ_u					1.33**		0,02**		0,04**	
σ_η	1,11**		1,11**		0.82**		0,83**			
λ_0									0,25**	
$\lambda_{D_{t-1}}$									0,46**	
$\lambda_{ln(M_{t-1})}$									-0,14**	