Direction des Études et Synthèses Économiques

G 2010 / 02

Health Expenditure Models: a Comparison of five Specifications Using Panel Data

Valérie ALBOUY - Laurent DAVEZIES
Thierry DEBRAND

Document de travail

Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des documents de travail
de la Direction des Études et Synthèses Économiques

G 2010 / 02

Health Expenditure Models: a Comparison of five Specifications Using Panel Data

Valérie ALBOUY* - Laurent DAVEZIES**
Thierry DEBRAND***

FÉVRIER 2010

The authors thank Brigitte Dormont, the participants of the INSEE-D3E seminar, Francesco Moscone and Alberto Holly for their support and Badi Baltagi, Jean-Paul Chaze, William Greene, Xavier d'Haultfoeuille, Andrew Jones, the participants of the Health Econometrics Workshop organized by the University of Leicester and the University of Milan-Bicocca and two anonymous referees for their helpful comments. We also thank Nelly Le Guen, Stephanie Guillaume and the entire ESPS team at IRDES. The usual disclaimer applies.

* Direction des Statistiques Démographiques et Sociales - Division des Études Sociales - Timbre F120 - 18 bd Adolphe Pinard - 75675 PARIS CEDEX 14

** Département des Études Économiques d'Ensemble - Division « »
Timbre G2 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

*** IRDES - 10 rue Vauvenargues - 75018 Paris

Département des Études Économiques d'Ensemble - Timbre G201 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX - France - Tél. : 33 (1) 41 17 60 68 - Fax : 33 (1) 41 17 60 45 - CEDEX - E-mail : d3e-dg@insee.fr - Site Web INSEE : http://www.insee.fr

Health Expenditure Models:
a Comparison of five Specifications Using Panel Data

Abstract

In this article, we focus on the estimation of outpatient expenditures with panel data. We model the logarithm of expenditures and consider five different models. The first two models are cross section two part and sample selection models. The two-part approach appears inappropriate when moving to panel data. We therefore focus on panel data models with sample selection. Our third model is a model without lagged dependent variables, and the last two ones include such lagged variables. These two latter models differ in their assumptions on the variance of the residuals. Modeling heteroskedasticity may indeed be important to avoid the bias due to the retransformation problem. We show that lagged dependent variables are important factors of heteroskedasticity. For the models with state dependence we provide a new solution to the initial conditions problem by using generalized residuals. We establish that panel data models highly improve the correlation explained by the model in the time-series dimension without damaging the fit in the cross-section dimension. For all indicators of fit, the model with state dependence and heteroskedasticity seems to dominate the others.

Keywords: Health econometrics, health expenditures, panel data, selection models, state dependence

---

Les modèles de dépenses de santé : une comparaison de cinq spécifications sur données de panel

Résumé

Dans ce travail, nous comparons cinq modélisations du logarithme de la dépense de soins ambulatoires. L'estimation de ces différents modèles repose sur des données de panel. Le premier groupe adopte une approche transversale avec, d'une part, un modèle de sélection en coupe et d'autre part un « two part model » - qui se révèle inapproprié lorsqu'on dispose de données de panel. Le second groupe est constitué de modèles de sélection longitudinaux : le premier d'entre eux n'intègre pas de variable de dépense retardée dans les explicatives à les différences des deux autres. Les deux derniers se distinguent par leur hypothèse sur la variance des résidus. La modélisation de l'hétéroscédasticité peut en effet avoir une importance cruciale pour réduire les biais dus au problème de retransformation. Nous montrons que les variables dépendantes retardées sont d'importants facteurs d'hétéroscédasticité. Pour les modèles avec dépendance d'état, nous proposons une nouvelle solution pour traiter du problème des conditions initiales en utilisant des résidus généralisés. Nous établissons que les modèles en panel améliorent fortement la modélisation de la corrélation des variables dépendantes dans la dimension temporelle sans pour autant dégrader l'ajustement en coupe. Pour tous les indicateurs d'ajustement considérés, le modèle intégrant dépendance d'état et hétéroscédasticité semble dominer les autres.

Mots-clés : Économétrie de la santé, dépenses de santé, données de panel, modèle de sélection, dépendance d'état

Classification JEL : I0, C1, C5

1 Introduction

In this paper, we focus on the estimation of equations on the medical consumption in a panel data framework. Such a work can be useful to better understand the dynamics of consumption over the lifecycle and to impute paths of health expenditures in a dynamic microsimulation model such as DESTINIE II.

Behaviour about health and medical care consumption is often conceptualised as a dynamic process (Jones, 2000; Jones, Rice and Contoyannis, 2006). As early as 1972, Grossman describes health as one 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. Medical reasons point also to the need to analyse health expenditures as a dynamic process. Indeed, illnesses may persist for a long time or have long-term consequences. A good understanding of health status and individual healthcare expenditure patterns requires taking past individual histories into account. In econometric terms, state dependence should be an important issue.

However, few studies focus on the dynamics of health expenditures. The main methodological articles on expenditure models with cross-section data focus on two main issues.

First, health economists have wondered whether or not it is necessary to model the decision-making process leading to the amount of care observed as a joint decision-making process (i.e. the decision to use on the one hand and how much to spend on the other) (Duan et al., 1983; Manning, Duan and Rogers, 1987; Hay, Leu and Rohrer, 1987; Hay and Olsen, 1984; Maddala, 1985). We establish that with panel data models, simple tests can be easily implemented to address this question. Our conclusion is that a Sample Selection Model must be used instead of a Two Part Model.

Second, the distribution of health expenditures is highly skewed and has heavy tails. Economists have often used concave transformations (logarithm or Box-Cox transformations) (Duan, 1983; Chaze, 2005) or generalised linear models to estimate health expenditures (see Manning and Mullahy, 2001 for an extensive discussion on the advantages and drawbacks of such models). Following a standard practice, we work in "log", assume the normality of residuals, and make standard assumptions on the covariance matrix of residuals for the panel models. The motivation of this choice is the need to have a simple parametric model when we consider the state dependence. The drawbacks of such an approach have been widely reported in cross-section analyses: Duan (1983) have proposed a non-parametric smearing estimator to relax non-normality but maintain the independence between residuals and covariates; Manning (1998), Mullahy (1998) and Manning and Mullahy (2001) among others point out that all of the moments of the residuals given covariates have an incidence on the mean prediction of the conditional expenditure.

For the models with state dependence, we provide a new solution to the initial condition problem by controlling for generalised residuals. We also show that lagged dependent variables are important factors of heteroscedasticity. As a result, in the most complete model, we consider the modelling of the conditional variance of residuals to avoid this problem. This modelling of variance is based on the use of lagged dependent variables, which is an extra motivation to use panel data.

The main results of this paper are the following. Panel data models highly improve the correlation explained by the model in the time-series dimension without damaging the fit in the cross-section dimension. For all indicators of fit, the model with state dependence and heteroscedasticity seems to dominate the others. This gives us guide-lines to implement paths of health expenditures in a longitudinal microsimulation model such as Destinie, because distribution (across individuals) of the net present value of transfer depends on the structure of the expenditures in the both dimension (cross-section and time-series dimension).

This paper is divided into three parts. In the second section, we present and discuss the advantages and drawbacks of five models. The first model is a two-part model in cross-section, and the second model is a sample selection model in cross-section. Models 3 to 5 are panel data models. In the Model 4, we introduce state dependence, and in Model 5 we use lagged dependent variables to deal with heteroscedasticity and the retransformation problem. In the third section, we present dataset, which merges administrative datasets from public health insurance and information from a survey on social welfare and health with many socioeconomic covariates. There is more than 7 000 individuals in the sample, and we observe their ambulatory expenditures from 2000 to 2005. In the fourth section, we discuss the estimation and the quality of the models. In particular, we compare the capacity of the models to fit the data in the cross-sectional and time-series dimensions.

2 The models

In this section, we present the five models and the reasons for the choice of the specifications.

2.1 Model 1: A cross-section two-part model

The distribution of health expenditures contains a high proportion of observation without consumption: in France, over a one-year period, 7% have no use of ambulatory care. Health economists have adopted several strategies for dealing with this mass point. A common framework can be adopted for all of the models we consider. It is made up of two equations; the first determines whether the individual has positive health care expenditures, and the second determines its amount (in the case of positive expenditure). We denote as D_{it} the fact that individual i has use at date t ($D_{it} = 1$ in the case of use and $D_{it} = 0$ otherwise). We note the amount of expenditure as M_{it}. When there is no ambiguity, indices i and t are omitted. We note as X^D the covariates 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 $X^D \neq X^M$ and for the empirical part of this paper $X^D = X^M = X$. The expenditure is deduced from the data (D_{it}, M_{it}):

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

$$ln(M) = (X^M\beta + \eta)D \tag{2}$$

The two-part model specification (2PM) assumes that $E(\eta|D=1,X) = 0$. It implies that the second equation models only $M|X, D=1$ and not a latent distribution $M|X$ only observed when $D = 1$. The researchers of RAND (Duan et al., 1983; Manning, Duan and Rogers, 1987) have promoted the use of such two part models.

2.2 Model 2: A sample cross-section selection model

Some researchers (Hay, Leu and Rohrer, 1987; Hay and Olsen, 1984; Maddala, 1985) support another specification: the sample selection model (SSM). The main argument in favor of the SSM is that common factors could simultaneously affect health care use and the amount of care. 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 classically made:

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} | X \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix}\right) \tag{3}$$

The coefficient β of SSM and 2PM cannot be compared, as these two models do not assess the same underlying economic model. In the 2PM, the coefficient β cannot be interpreted as the causal impact of X on amount M (because individual heterogeneity influences participation D).

A SSM model should be used in order to estimate the effect of the variables X on the amount of health expenditure, all other things being equal (including health status, which is always imperfectly observed and which is therefore included in ε and η). 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 does not matter if the purpose is to predict expenditures. This is the argument highlighted by Manning, Duan and Rogers in 1987.

2.3 Model 3: A panel data sample selection model

Assumptions on the unobserved heterogeneity in equations (1) and (2) can be relaxed with panel data: in a first time, we only assume that $(\varepsilon_{it}, \eta_{it})_{t=1,...,T}$ are independent across individuals and we discuss the impact of this assumption on the joint distributions of $(\varepsilon_{it}, \eta_{it})_{t=1,...,T}$ for each individual.

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

A two part model specification with panel data assumes that $(\varepsilon_{it})_{t=1,...,T} \perp\!\!\!\perp (\eta_{it})_{t=1,...,T}|X$. The following proposition gives some testable implications of this assumption.

Proposition 1:
Under a data generating process described by equations (1) and (2), we have:

$$(\varepsilon_{it})_{t=1,...,T} \perp\!\!\!\perp (\eta_{it})_{t=1,...,T}|X \Rightarrow M_{i\tau} \perp\!\!\!\perp (D_{it})_{t=1,...,T}|(X, D_{i\tau}).$$

Indeed, $ln(M_{i\tau})$ is a function of $(X, \eta_{i\tau}, D_{i\tau})$, thus $ln(M_{i\tau}) \perp\!\!\!\perp (\varepsilon_{it})_{t=1,...,T}|X, D_{i\tau}$. Because $(D_{it})_{t=1,...,T}$ is a function of $(X, (\varepsilon_{it})_{t=1,...,T})$, the result follows.

This implies many testable restrictions such as $M_{i\tau} \perp\!\!\!\perp \sum_{t=1}^{T} D_{it}|X, D_{i\tau} = 1$ or $E\left(M_{i\tau}|X, D_{i\tau} = 1, g((D_{it})_{t=1,...,T})\right) = E\left(M_{i\tau}|X, D_{i\tau} = 1\right)$ for any function g. Moreover the tests of these restrictions do not rely on parametric assumptions such as normality or on exclusion restrictions, unlike the test of a sample selection model versus a two part model estimated with cross-section data.

Note that proposition 1 holds whatever the structure of autocorrelation of ε_{it} and η_{it}: we make no assumption about $cov(\varepsilon_{it}, \varepsilon_{it'})$ and $cov(\eta_{it}, \eta_{it'})$. It does not rely either on independence between unobserved heterogeneity and the covariates, so proposition 1 includes the case of fixed effect panel data models.

Testable restrictions obtained with the proposition 1 are rejected as soon as some unobserved variables constant in the times-series dimension have a simultaneous effect on the use and on the amount of expenditures. This is the case in our data, as the consumption per year of individuals who often use health care is much higher than that of other individuals (see for instance Table 6 in 3.2). To test this correlation given the observables X, we regress M_{it} on $\sum_{t=1}^{T} D_{it}$ and X (respectively on $\sum_{t=1}^{T} D_{it}$, $\left(\sum_{t=1}^{T} D_{it}\right)^2$ and X) conditionally on $D_{it} = 1$. The Fisher statistic is equal to 314 (respectively to 183) and in both case, the p-value are lower than 1/10 000. This leads us to prefer an SSM model rather than a 2PM model in a panel framework.

With T periods, the individual matrix of variance-covariance depends on $T(2T+1)$ terms and the likelihood is an integral[1] on $\mathbb{R}^{2T}$. For the sake of simplicity, even if such restrictions are not necessary for the identification and the estimation of the model, we assume for each equation (use and amount) that the unobserved heterogeneity can be divided into a time invariant component and an idiosyncratic one, following a widely used framework. So, Equations (1) and (2) become:

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

u_i and v_i can be interpreted as follows: some individuals with the same characteristics X use treatment more frequently than others, and among individuals who 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 the frequency of health care use and of the amounts observed per individual. Contrary to cross-section data, these frequencies are observable with panel data. We assume that pairs (u_i, v_i) are independent and identically distributed. Moreover, we assume that u_i and v_i are independent of X. In other terms we use a random effect panel data model. We could have used a fixed effect model without any assumption on the joint distribution of (u_i, v_i, X). Nevertheless this would raise estimation difficulties linked to the problem of incidental parameters[2]. Kyriazidou (1997) proposed a method to estimate consistently β and γ by using observations such that $X_{it}^{D}\gamma = X_{it'}^{D}\gamma$. Thus, estimation relies only on individuals for which we can find others covariates that balance the effects of age. So, this method is not appropriate when considering the substantial effects of age.

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

Following the test based on proposition 1, we do not assume that $u_i \perp\!\!\!\perp v_i|X$ or that $\varepsilon_{it} \perp\!\!\!\perp \eta_{it}|X$

In the case of selection models and in the absence of a large support instrument, it is usual (and necessary) to make parametric assumptions concerning the distribution of unobserved heterogeneity. To estimate the model we assume[3] that:

[1] To reduce the dimensionality of the integral, β and γ can be estimated by Quasi-Maximum-Likelihood or by two step estimation, using inverse Mill's ratio but these methods are not valid with state dependence that we introduce in the two last models we estimate.

[2] Using individual fixed effects increase the dimensionality of the nuisance parameters with the size of sample. In this case, maximum-likelihood estimators have non-standard properties (for non linear models).

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

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

When reasoning for a period and for an individual, the likelihood of the observation $(D_{it}, ln(M_{it}))$ is therefore
$\mathcal{L}_{ln(M_{it})|X_{it},u_i,v_i}(ln(M_{it}))P(D_{it}=1|X_{it},u_i,v_i,ln(M_{it}))D_{it} + P(D_{it}=0|X_{it},u_i,v_i)(1-D_{it})$.

The likelihood of an individual can then be deduced easily:

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

with

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

and

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

2.4 Model 4: Introducing state dependence

Although many authors work on panel datasets, to our knowledge very few papers deal with dynamic health models. Noland (2007) analysed the dynamics of general practitioner visits; Contoyannis, Jones and Rice (2004) looked at the dynamics of health status. Deb, Trivedi and Zimmer (2009) propose a dynamic two-part model for health care expenditures. However, we are not aware of any paper that estimates a sample selection model of health expenditures with state dependence.

Our model is written in the following way[4] for date $t > 0$:

$$D_{it} = \mathbf{1}_{\{D_{it-1}\gamma_1 + ln(M_{it-1})\gamma_2 + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}} \tag{4}$$
$$ln(M_{it}) = (D_{it-1}\beta_1 + ln(M_{it-1})\beta_2 + X_{it}\beta + v_i + \eta_{it})D_{it} \tag{5}$$

We face a main difficulty when estimating coefficients $\gamma_1, \gamma_2, \gamma, \beta_1, \beta_2, \beta$ since $(D_{it-1}, ln(M_{it-1}))$ is correlated with (u_i, v_i). Because of the markovian properties of the model, we have:

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

Equations (4) et (5) imply that (u_i, v_i) and initial conditions $D_{i0}, ln(M_{i0})$ are correlated (given X). Therefore it is not possible to write the joint distribution $\mathcal{L}_{D_0,ln(M_0),u,v|X}$ as a product of marginal distributions:

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

[4] As for the left hand side, we use the convention $log(0) = 0$ for the right hand side of the equations 4 and 5. Thus if $D_{it-1} = 0$ then $log(M_{it-1}) = 0$.

This problem reflects the difficulty of disentangling state dependence and individual heterogeneity.

Heckman (1981) studied the problem of endogeneity of initial conditions in a binary model with state dependence. If we extrapolate this method, the aim is to approach the distribution $D_0, ln(M_0)|u, v, X$ by additional parametric assumptions. The first problem is that these supplementary assumptions are not consistent with the assumptions on the distribution of $D_{it}, ln(M_{it})|D_{it-1}, ln(M_{it-1}), X, u, v$. So the convergence of the estimators is ensured only when $T \rightarrow +\infty$, that is in the case when the initial condition problem becomes trifling. In his works on the binary model, Heckman highlighted from simulations Monte Carlo that the estimators can be strongly biased in the case of few periods. In our case, we have a panel of 6 periods thus this method seems inappropriate.

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

$$(u_i, v_i)|D_{i0}, ln(M_{i0}), X_i \sim \mathcal{N}\left(D_{i0}K_D + ln(M_{i0})K_M + X_iK_X, \Sigma\right)$$

This approach is very close to that of Chamberlain for getting round the problem of incidental parameters in the case of non linear panel data models. The heterogeneity is divided into an observable part which depends on X and initial conditions (the term $D_{i0}K_D + ln(M_{i0})K_M + X_iK_X$) and an unobservable part modeled by a normal random effect.

The likelihood can be divided into two parts:

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

Under the additional assumptions of Wooldridge, it is possible to estimate the following model for $t > 0$ on the one hand:

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

and to estimate the distribution of $D_0, ln(M_0)|X$ in cross-section on the other hand.

One of the main advantages of Wooldridge's method lies in its ease of implementation. Variables $D_0, ln(M_0)$ and $(X_t)_{t=0...T}$ play exactly the same role in the estimations as covariables X, and the usual estimation techniques apply.

However, the number of parameters becomes huge (the size of K_X is $K \times T$). Moreover collinearity or quasi-collinearity problems may occur between X and $(X_t)_{t=0...T}$ and between $D_{t-1}, ln(M_{t-1})$ and $D_0, ln(M_0)$. First, we have to assume that $\kappa_{u,X_{it}}$ and $\kappa_{v,X_{it}}$ are null for covariates without variations in the time series dimension. We can make some further restrictions on K_X assuming that $\kappa_{u,X_{it}}$ and $\kappa_{v,X_{it}}$ equal zero for $t > 0$ or for $t \geq 0$ for covariates with weak variation in the time series dimension. But these restrictions can lead to biased estimatimates of β and γ, when they fail to hold.

Regarding our data, it was impossible to separate the different effects convincingly. Correlation of age and consumption at date $t = 0$ with age and consumption at the following dates cause very unstable estimation without restrictions on K_X. If we use restrictions on $\kappa_{u,X_{it}}$ and $\kappa_{v,X_{it}}$ for $t > 0$, by including only D_{i0}, $ln(M_{i0})$ and X_{i0} as supplementary covariates, estimations remain unstable. And if we assume such restrictions for all t, by including only D_{i0}, $ln(M_{i0})$ as supplementary covariates, estimations become clearly unconvincing (especially for the coefficients of age and the coefficients of the previous state). To get round these difficulties of quasi-collinearity, we propose an adaptation of Wooldridge's method.

As Wooldridge did, we assume that the heterogeneity takes the form:

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

Wooldridge restricts f to the set of linear functions. This set is therefore quite large if there are a lot of variables X and a lot of periods. Consequently, we propose reformulating Wooldridge's assumption by considering a more limited set of functions f. $f(D_0, ln(M_0), X)$ models the over-propensity (respectively under-propensity) of using health care and the over-propensity (respectively under-propensity) of having high expenditures. We therefore propose using the generalized residual r_0 of a simple model in cross-section at the initial date. The generalized residual is the best available estimation of the over or under propensity to consume at the initial date. In a linear framework, the generalized residual 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 $E\left(\xi|Y_0, X_0\right)$.

We used a 2PM in $t = 0$ to estimate the generalized residuals. In a 2PM, the generalized residual satisfies:

- For a consumer in t=0:

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

- For a non consumer in t=0:

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

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

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

2.5 Model 5: Introducing heteroskedasticity

We now take into account the possible heteroskedasticity of residuals because of the retransformation problem. The descriptive statistics highlight strong heteroskedasticity in the logarithm of the amount: indeed the scattergram representing the logarithms of the amounts of two successive consumptions is considerably flattened for large amounts (see Figures 1 and 2). In other terms, it seems that $V(ln(M_{it})|ln(M_{it-1}), D_{it-1} = 1)$ decreases with $ln(M_{it-1})$. On the other hand, $V(ln(M_{it})|X_{it}, D_{it} = 1)$ seems to be constant when we plot the graph $(X_{it}, ln(M_{it}))$ with age or other covariates. To be parsimonious, we assume that $V(ln(M_{it})|X_{it}, ln(M_{it-1}), D_{it-1} = 1) = V(ln(M_{it})|ln(M_{it-1}), D_{it-1} = 1)$.

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

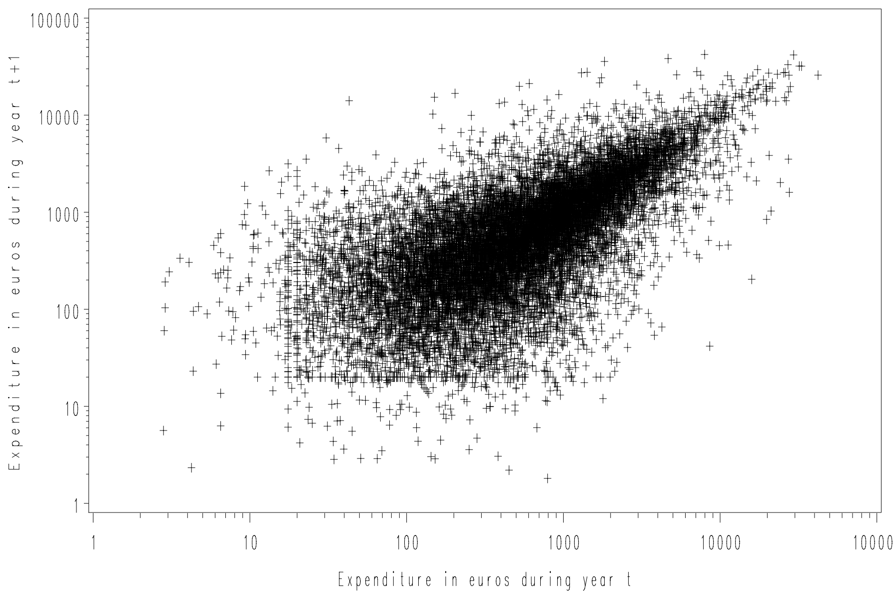
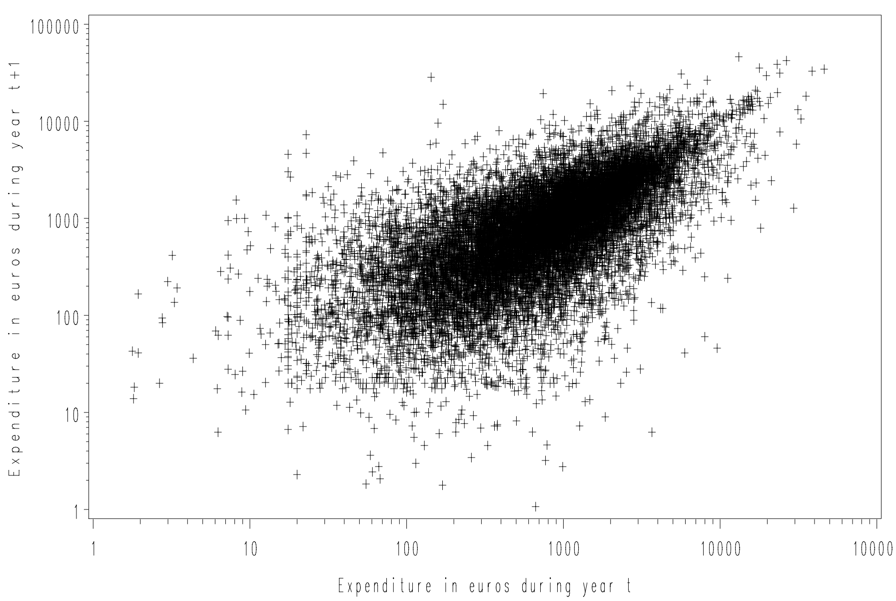


Figure 2: Expenditure in t and $t+1$ for women (logarithmic scale)



The omission of this heteroskedasticity is liable to bias the estimation of coefficients β and γ (contrary to the linear model or to the two part model), and it also raises retransformation problems, since the average level of the amounts is related to the dispersion of log-amounts[5]. It is therefore necessary to model possible heteroskedasticity. This is what we have done by assuming that:

[5] In the gaussian case, the problem of retransformation is linked to the second order moment of $\eta|ln(M_{it-1}), D_{it-1}=1, X_{it}$ because $E(e^{\eta}|ln(M_{it-1}), D_{it-1}=1, X_{it}) = e^{\frac{V(\eta|ln(M_{it-1}),D_{it-1}=1,X_{it})}{2}}$

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

The model with heteroskedasticity becomes:

$$D_{it} = \mathbf{1}_{\{D_{it-1}\gamma_1 + ln(M_{it-1})\gamma_2 + r_{D_{i0}}\gamma_{r_{D_0}} + r_{ln(M_{i0})}\gamma_{r_{ln(M_0)}} + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}}$$
$$ln(M_{it}) = \left(D_{it-1}\beta_1 + ln(M_{it-1})\beta_2 + r_{D_{i0}}\beta_{r_{D_0}} + r_{ln(M_{i0})}\beta_{r_{ln(M_0)}} + X_{it}\beta + v_i + \eta_{it}\right) D_{it}$$

The conditional 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 $r_{ln(M_{i0})} = E(\tilde{\varepsilon}_{i0}|D_{i0}, ln(M_{i0}), X_{i0})$ and $r_{D_{i0}} = E(\tilde{\eta}_{i0}|D_{i0}, ln(M_{i0}), X_{i0})$ can be estimated by root-N convergent estimators of the generalized residuals of the following two part model:

$$D_{i0} = \mathbf{1}_{\{X_{i0}\gamma_0 + \tilde{\varepsilon}_{i0} > 0\}}$$
$$ln(M_{i0}) = (X_{it}\beta_0 + \tilde{\eta}_{i0}) D_{i0}$$

3 Data and descriptive statistics

3.1 The specificities of ESPS – EPAS

The empirical work uses matched administrative and survey data. The survey data is based on Social Welfare and Health Surveys (ESPS) from the IRDES[6]. ESPS provides detailed information on socioeconomic situation and health status (disability, self reported health, health status index). We matched these survey datasets with administrative datasets from public national insurance (Public health insurance for employees): the Permanent Health Insurance Samples (EPAS). EPAS provides detailed information relative to health expenditures. For every person in the ESPS 2000 dataset, we collected health expenditures (amount and use) from EPAS for the period 2000-2005. This provides a panel dataset over 2000-2005 with 7,085 individuals. This panel is unbalanced due to several reasons. In very particular cases, individuals may have moved out of the main health insurance system (for instance civil servants have a different Public health insurance system), or have died. For each year, in the EPAS, people who died are omitted on the basis of informations given by the registry office; moreover we used the ESPS 2004 survey which is available for the half sample to check the dataset quality concerning the death (see Table 2).

Our panel data is quite unusual in the sense that socioeconomic data are only available in 2000, while health expenditures are available from 2000 to 2006 (see Table 1). Consequently, all the characteristics in our model are time invariant (except age and time to death), which is problematic when disentangling them with constant unobserved heterogeneity.

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

Table 1: Panel dataset

	2000 ESPS-EPAS	2001-2003 EPAS	2004 ESPS-EPAS	2005 EPAS
Socioeconomic characteristics	Age Gender Employment status Education level Household structure		reasons of exit of the scope of EPAS (for half sample)	
Health expenditures	Ambulatory use and amount	Ambulatory use and amount	Ambulatory use and amount	Ambulatory use and amount

Note : in 2004, for the half sample we have some indications about attrition: death, moving abroad, change to a special social security regime not covered by the EPAS datasets (for very specific socio-professional groups like mineworkers, representing less than 5% of the French population).
Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005).

Table 2: Descriptive Statistics

	Nb of Indiv.	use D	amounts in € $M\|D=1$	Nb of deaths
2000	7 085	0.92	1 013	60
2001	7 025	0.93	1 083	43
2002	6 982	0.95	1 173	58
2003	6 924	0.95	1 231	72
2004	6 852	0.92	1 325	45
2005	6 807	0.90	1 363	65
All	41 675	0.93	1 195	343

Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005).

3.2 Descriptive statistics

Tables 2 to 5 present simple statistics: 92% of the individuals of our panel had a use in 2000. Among them, the average expenditures in 2000 is €1,013. Concerning socio economic characteristics, 37% of adults are employee (non managerial), 21% retired, 10% executive and 10% public servant, 8% unemployed, and 11% are not retired but out of the labor market and a small part (less than 2%) is self-employed and has significantly lower ambulatory expenditure than others. Unemployed people have a relatively low probability (88 %) to use an ambulatory care but when they have a positive expenditure, it is larger than the average expenditure of other workers (€1,144 versus €853 for employees). Significant differences in the level of expenditures exist between more educated and less educated (less than €1,000 for higher level of education versus more than €1,700 for primary school level of education). When they have positive expenditures, the 911 children of employees in the ESPS-EPAS panel have lower expenditures (€417) than the 318 children of executives (€500). The postsecondary students who live with their parents have a low probability of positive expenditures (0.33 %), which could be partially due to the fact that children have a worse health coverage when they come of age. Difference of expenditures with the level of education can be due to age effect: indeed, the younger generations are more educated than the older and the consumption increases with age (Table 4).

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 individual healthcare expenditure. Our data confirm this fact: the variability of health care use is huge between individuals who will die in the next year (who spend €4,404 in ambulatory care) and those who will live more than two years afterwards (who spend €1,153).

These descriptive statistics show great variability in levels of health expenditure: the first quartile of positive expenditures is €211 and the third is €1,204. The right tail is heavy (Table 5). 55.0% of men and 41.1% of women spend more than €500 in a year whereas only 3.1% of men and women spend more than €5,000 per year, and 0.3% of the individuals spend more than €15,000 in a year for outpatient care. If we select only individuals alive at the end of 2005 and we 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 with their frequency: the median of the amounts of consumption is greater than €700 for those who had consumed every year versus less than €150 for those who consumed less than one year in two (Table 6).

3.2.1 Covariates

We selected the following covariates for our estimations: (i) AGE: a polynomial of order 4 for age; (ii) TIME TO DEATH: dummy variables, one and two years before death. This allows to take into account the acceleration of health expenditure before death. These two variables are interacted with age; for others covariates, we differentiate adults from young people who live with their parents. For adults we add: (iii) EMPLOYMENT STATUS: a set of dummy variables which only concern adults who are self-employed, private sector managers, civil servants, unemployed, pensioners, other inactive persons. The reference modality is private sector employees (non managerial). These variables are put to zero for young people who live with their parents; (iv) LEVEL OF EDUCATION: a set of dummy variables which corresponds to the highest level of education obtained

Table 3: Health expenditures and characteristics in 2000

	Nb	D	ambulatory $M\|D = 1$(in €) Mean	Q1	Q3
Gender					
Male	3 447	0.90	922	165	1 019
Female	3 638	0.94	1 095	269	1 328
Adult	5 327	0.93	1 197	277	1 439
Employment status					
Self-employed	92	0.71	549	120	700
Executive	551	0.95	934	233	1 051
Employee	2 013	0.92	853	200	1 027
Public service	527	0.92	947	237	1 136
Unemployed	412	0.88	1 144	222	1 367
Retired	1 128	0.97	1 997	776	2 452
Other non-working	604	0.95	1 353	395	1 624
Level of education					
primary school	953	0.95	1 710	540	2 123
school	2 051	0.93	1 054	237	1 234
secondary school	813	0.93	1 129	281	1 423
advanced	1 187	0.93	993	227	1 182
other schooling career	141	0.95	1 416	494	1 886
no schooling	182	0.92	1 535	396	1 854
Young	1 758	0.90	433	124	547
Father's status					
Self-employed	119	0.90	391	106	514
Executive	318	0.91	500	160	599
Employee	911	0.92	417	116	532
Public service	187	0.91	464	135	620
Unemployed	110	0.86	331	124	459
Retired	28	0.64	295	113	552
Other non-working	85	0.76	514	91	574
Level of education					
primary school	866	0.93	345	108	445
school	429	0.92	548	119	704
secondary school	117	0.86	501	172	655
advanced	82	0.33	329	76	356
other schooling career	9	1.00	1 569	308	2 331
no schooling	255	0.95	479	221	589
All	7 085	0.92	1013	211	1204

Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005). Computation using all individuals and only the first period.

Table 4: Health expenditures, age and time to death (all periods)

	Nb	ambulatory			
		D	$M\|D=1$(in €)		
			Mean	Q1	Q3
Age					
0-29	14 171	0.89	537	133	659
30-44	10 694	0.93	872	210	1 035
45-59	8 920	0.94	1 455	372	1 698
60-74	5 121	0.97	2 131	765	2 561
75 +	2 494	0.99	2 952	1 242	3 490
Time to death					
Time to death > 2 years	41 087	0.93	1 153	231	1 374
1 year < Time to death ≤ 2 years	266	0.95	3 620	780	4 498
Time to death ≤ 1 years	322	0.96	4 404	1 044	5 299
All	41 675	0.93	1 195	233	1 405

Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005).
Computation using all individuals and all periods.

Table 5: Ambulatory expenditures by gender (€)

	Men		Women		All	
	%	M(€)	%	M(€)	%	M(€)
no expenditure	8.8	0	5.6	0	7.2	0
0 - 499 €	46.2	207	35.5	231	40.7	218
500 € - 999 €	18.0	724	20.9	730	19.5	728
1 000 € - 4 999 €	23.9	2 044	34.9	2 023	29.6	2 031
4 999 € - 14 999 €	2.7	7 743	2.8	7 464	2.7	7 596
15 000 € +	0,4	21 635	0.4	20 469	0.3	21 110
All	100	1 004	100	1 209	100	1 110

Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005).
Computation using all individuals and all periods.

Table 6: Correlation between frequence and use

$\sum_{t=1}^{T} D_{it}$	0	1	2	3	4	5	6
N	4	39	83	213	478	771	5 219
%	0,06	0,57	1,22	3,13	7,02	11,33	76,67
$E\left(M_{it}\|D_{it}=1\right)$	-	531	304	360	499	492	1 294
$Med\left(M_{it}\|D_{it}=1\right)$	-	132	121	121	226	245	736

Source : ESPS Survey 2000 and EPAS datasets (2000 to 2005).
Computation on individuals alive in 2005

by adult: primary, Bachelor degree, higher, other education. The reference is a master's degree. These variables are put to zero for young people who live with his or her parents. For this people, we add: (v) YOUNG: a simple dummy (vi) EMPLOYMENT STATUS OF THE HEAD OF HOUSEHOLD: which is taken into account according to dummies mentioned above (vii) CURRENT LEVEL OF STUDIES: which is taken into account according to the dummies mentioned above. These variables are put to zero for adults. In other terms, the X are : a flexible function of AGE, TIME TO DEATH, TIME TO DEATH*AGE,(1-YOUNG)*(CONSTANT, EMPLOYMENT STATUS, LEVEL OF EDUCATION) and YOUNG*(CONSTANT, EMPLOYMENT STATUS OF FATHER, CURRENT LEVEL OF EDUCATION OF THE YOUNG).

4 Results

4.1 Estimation and factors of health expenditures

The results of the estimations are reported in Tables 11 and 12 in appendix. Health expenditure is correlated with age, "time to death" and social status (employment status, educational level) as already stated by many authors (Zweifel, Felder and Meiers, 1999; Zweifel, Felder and Werblow, 2004; Stearns and Norton, 2004; Seshamani and Gray, 2004a, 2004b).

Concerning age, the effects are globally significant and positive (see Tables 11 and 12). Concerning the time to death we find positive effects on the probability to use and on expenditures. As usual for the cross variable between age and time to death we obtain a negative effect: this could reflect that lethal pathologies vary with age.

Concerning employment status, self-employed people always have a lower probability of use and lower expenditures than others. This is a classical result that can be explained by the fact that self-employed are financially prompt to work even if they are ill (at the difference of others). Unemployed people have a lower probability to use a health care. We find the same results for the children of self-employed and those of unemployed. The children of retired people have also a lower probability to consume and lower expenditures when they use health care. The children of executives have higher expenditures than children of employees (non managerial).

The level of education has a very different effect for men and women: for instance, men with advanced level of schooling have higher probability of use than men with secondary school education, but we have the reverse for women. This difference may come from the fact that women with higher level of education have lower probability to have children and hence lower gynecologic expenditures.

Model 3 shows that unobserved heterogeneity is for a large part constant in the times series dimension : for the positive amount, σ_v and σ_η have approximatively the same magnitude for men and women, for the probability of use σ_u and σ_ε are closed for men but for women σ_u is greater than σ_ε.

State dependence (Models 4 and 5) is large. An increase of 1% of the amount at period $t-1$ induces a 0.25% increase of the amount at next period t for men and women. This effect is very likely due to the dynamics of the underliyng health status.

4.2 Comparison of models

The main objective of our paper is to compare the fit of each model. These models are nested: 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 ($\sigma_v = \sigma_u = 0$). 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. Because a variance cannot be negative, the test of $\sigma_u = \sigma_v = 0$ versus $\sigma_u > 0$ or $\sigma_v > 0$ is a test at boundary. We rely on Self and Liang (1987) procedure and reject the null hypothesis $\sigma_u = 0$ and $\sigma_v = 0$[7].

A second way to compare the models is to compare the capacity of the model in terms of prediction. For each individual, for each period (and for each model), we can compute a prediction of M and D on the panel dataset by drawing $\varepsilon_0, \eta_0, u, v, \varepsilon, \eta$ in the estimated distributions. Then we compare for different criteria the original data and the simulations for each model.

Adjustment for the probability of use can be measured by $\sum_{i=1}^{N}\sum_{t=1}^{T_i} |D_{it} - \widehat{D_{it}}| / \sum_{i=1}^{N} T_i$ which is the mean square error and the mean absolute error (because $\widehat{D_{it}} \in \{0,1\}$). For the amount of expenditures we compute the root mean square error ($\left[\sum_{i=1}^{N}\sum_{t=1}^{T_i}(M_{it} - \widehat{M_{it}})^2 / \sum_{i=1}^{N} T_i\right]^{\frac{1}{2}}$) and the mean absolute error($\sum_{i=1}^{N}\sum_{t=1}^{T_i} |M_{it} - \widehat{M_{it}}| / \sum_{i=1}^{N} T_i$). All of these indicators are some cross-sectional criteria. Theoretically, if the panel models are expected to better reproduce data in the time series dimension than cross-sectional models, in the cross-sectional dimension they could be worst. Results reported in Tables 7 and 8 show that the fitting of panel data models are equivalent to the cross-sectional models if we focus on use, and that the panel data model with state dependence and heteroscedasticity (Model 5) is better than the cross-sectional models if we focus on expenditures. The homoscedastic panel data models (Models 3 and 4) are equivalent to the cross-sectional model for men, but worst for women.

	Model 1	Model 2	Model 3	Model 4	Model 5
$\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i,t} \lvert D_{it} - \widehat{D_{it}}\rvert$	0.1530	0.1557	0.1538	0.1500	0.1537
$\left[\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i,t}(M_{it} - \widehat{M_{it}})^2\right]^{\frac{1}{2}}$	3552	3437	3617	3274	2991
$\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i,t} \lvert M_{it} - \widehat{M_{it}}\rvert$	1321	1300	1340	1304	1221

Note : $N = 3447$, $\sum_{i=1}^{N} T_i = 20173$

Table 7: Adjustment criterion between data and prediction, Men

The third criterion we retain is the capacity of models to reproduce the full distribution of the health expenditures. In the cross-sectional dimension, we compare the density of positive expenditures ($M_{it}|D_{it} = 1$) on the data and on the simulated expenditures for each model. The graphs are reported in Figure 3 and 4. For women (Figure 4) the five models seem to be approximatively equivalent. For men, density generated by Model 5 is

[7] More precisely, we test separately $\sigma_u = 0$ and $\sigma_v = 0$. A joint test is possible but harder to compute, the p-value of the two tests are sufficiently low to make highly improbable the acceptation of the null hypothesis $\sigma_u = \sigma_v = 0$ with a joint test.

	Model 1	Model 2	Model 3	Model 4	Model 5
$\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i,t} \lvert D_{it} - \widehat{D_{it}} \rvert$	0.1005	0.1003	0.1026	0.1038	0.1050
$\left[\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i,t} (M_{it} - \widehat{M_{it}})^2\right]^{\frac{1}{2}}$	3183	3207	3733	3504	2685
$\frac{1}{\sum_{i=1}^{N} T_i} \sum_{i,t} \lvert M_{it} - \widehat{M_{it}} \rvert$	1436	1404	1465	1472	1338

Note : $N = 3638$, $\sum_{i=1}^{N} T_i = 21502$

Table 8: Adjustment criterion between data and prediction, Women

slightly closer to the density estimated on data especially for the amounts between €50 and €1,200.

Age is a determinant factor of the health expenditures, so we compare the five models for their aptitude to mimic the proportion of use by age $E(D_{it}|age)$ and the mean amount of expenditures by age $E(M_{it}|D_{it} = 1, age)$. For the proportion of use (Figures 5 and 6), the five models seem roughly equivalent. We interpret the discontinuity in the probability of use at 20 or 21 by the fact that when young leave their parents they have to find a doctor near their new residence, by the fact that they change their Social Security cover and by the fact that their parents don't pay anymore for their consumption. The average amount by age for positive expenditures (Figures 7 and 8) shows that the five models have the same performance.

Last, to compare the models in the time series dimension, we compute the serial correlations: $Corr(D_{it}, D_{it-1})$, $Corr(D_{it}, M_{it-1})$, $Corr(M_{it}, D_{it-1})$ and $Corr(M_{it}, M_{it-1})$. Model 5 clearly outperforms the other models in the time-series dimension. Indeed, temporal correlations of health care amounts are stronger in Model 5 (see Tables 9 and 10), even if they remain underestimated compared to those observed in the data. Using panel data models and paying attention to the modelling of heteroscedasticity clearly improve the performance of the models in the time-series dimension.

	Data	Model 1	Model 2	Model 3	Model 4	Model 5
$CORR(D(t), D(t-1))$	*0.34927*	0.05297	0.03868	0.27915	0.40128	0.33328
$CORR(D(t), M(t-1))$	*0.12686*	0.04571	0.04556	0.08767	0.10347	0.10826
$CORR(M(t), D(t-1))$	*0.11728*	0.04346	0.04099	0.09037	0.10643	0.10292
$CORR(M(t), M(t-1))$	*0.71493*	0.11836	0.11978	0.40226	0.40848	0.61499

Table 9: Correlation over the time of ambulatory expenditure, Men

5 Conclusion

To conclude, a main difficulty in the study of paths of healthcare consumption is to account for the correlation of behavior over time. This paper highlights several implications:

First, if we are interested in the quality of estimations in the time-series dimension, it is important to take the selection into account. Indeed, the correlation between amounts of expenditure in case of use and the frequency of use can not be explained only by observable

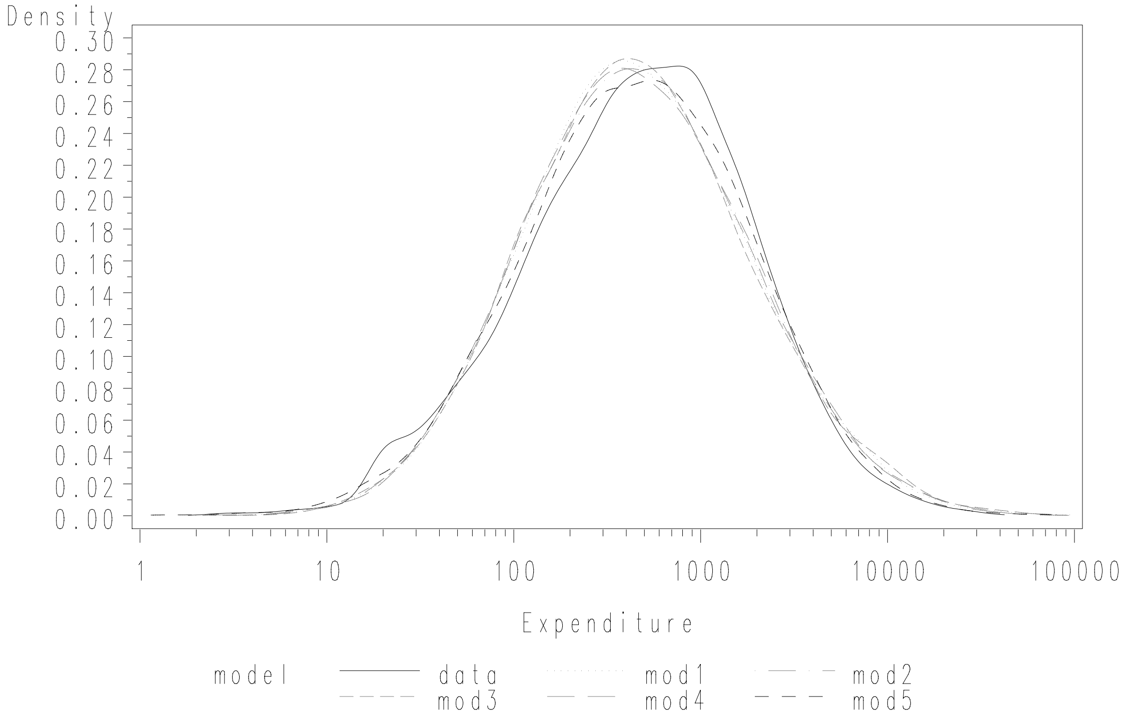


Figure 3: Density of amounts for men (in €)

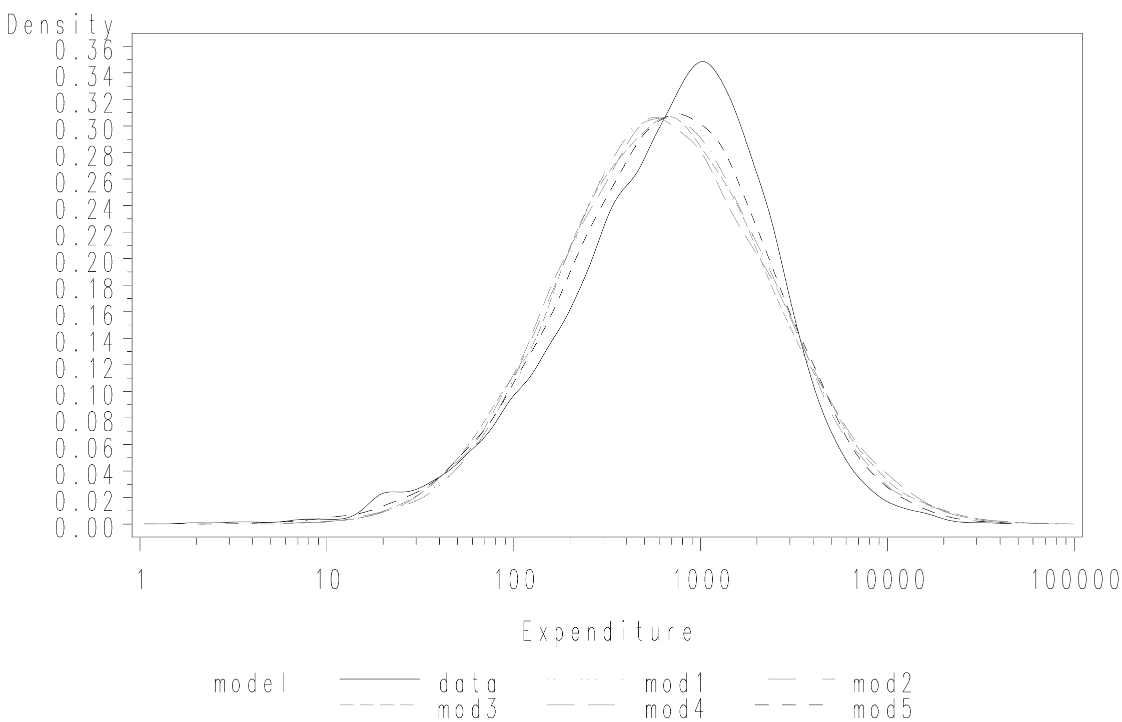


Figure 4: Density of amounts for women (in €)

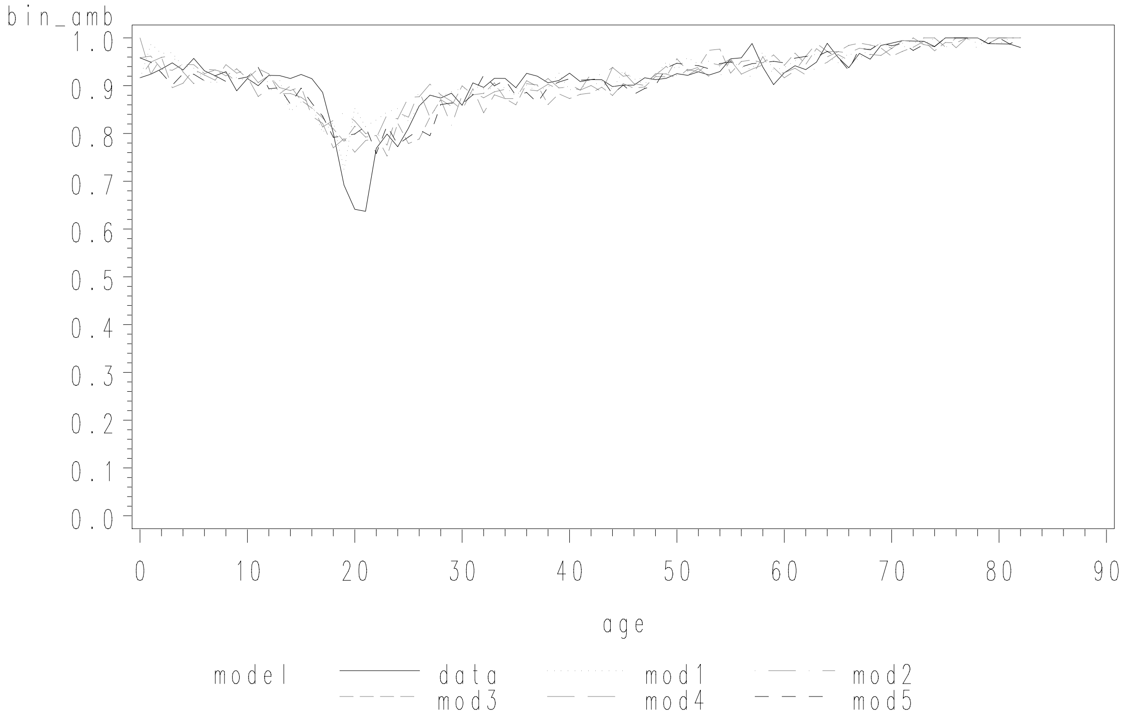


Figure 5: Probability of use by age for men

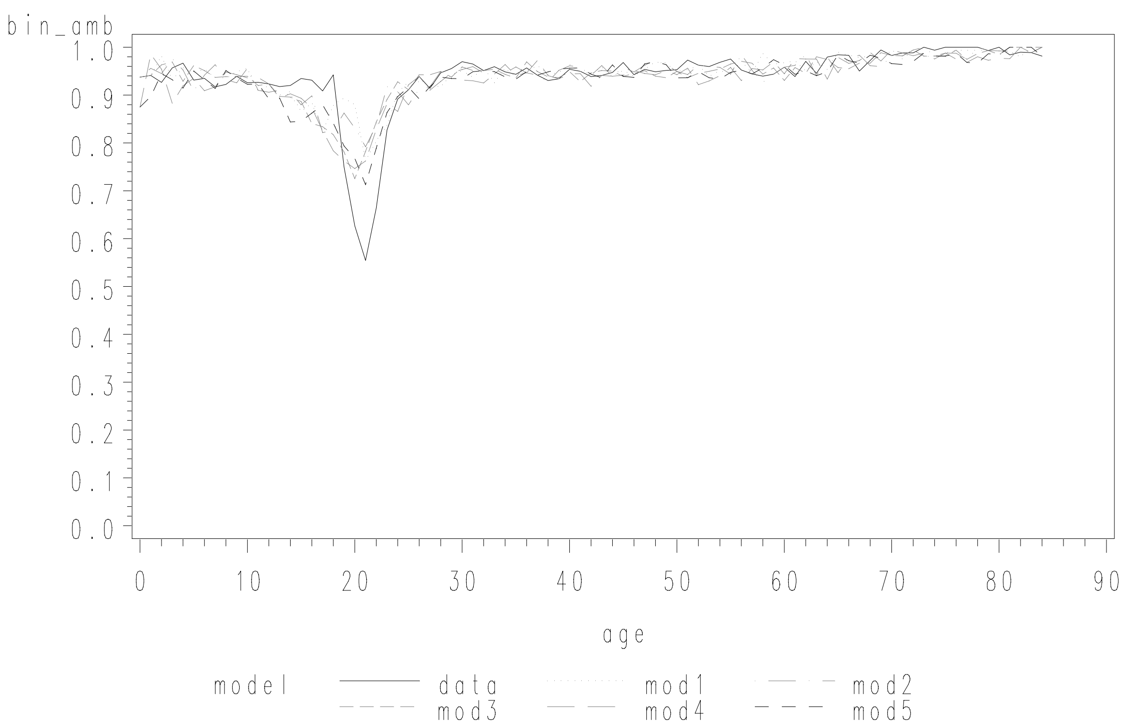


Figure 6: Probability of use by age for women

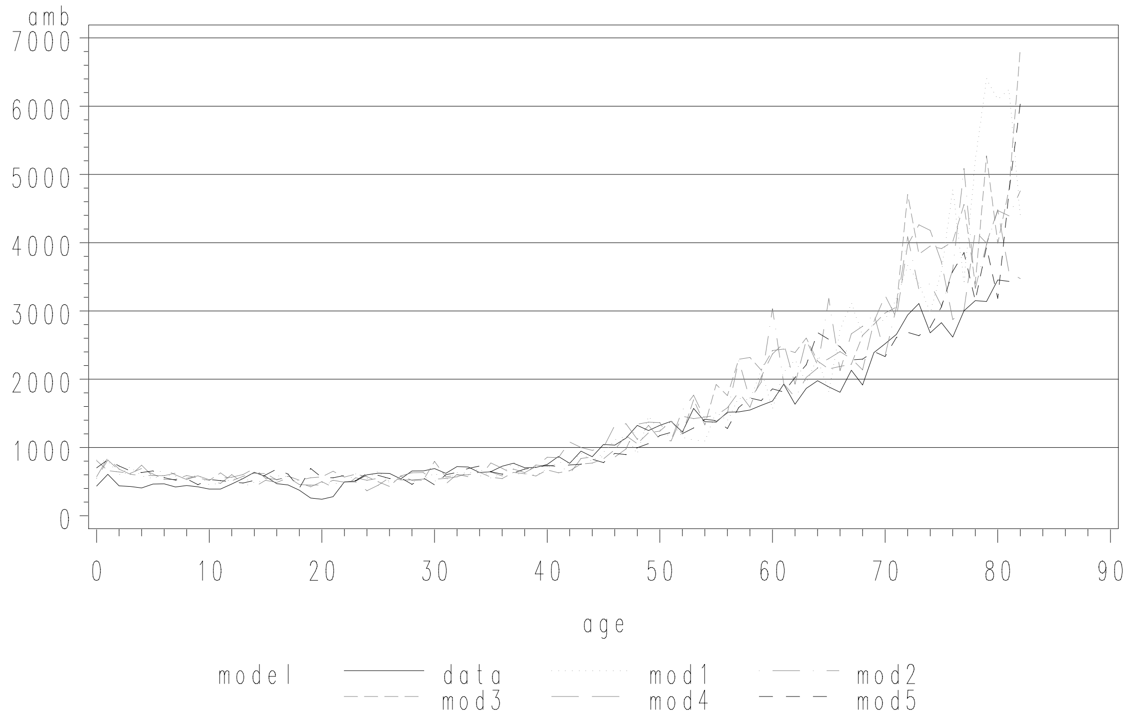


Figure 7: Average of expenditure by age for men (in €)

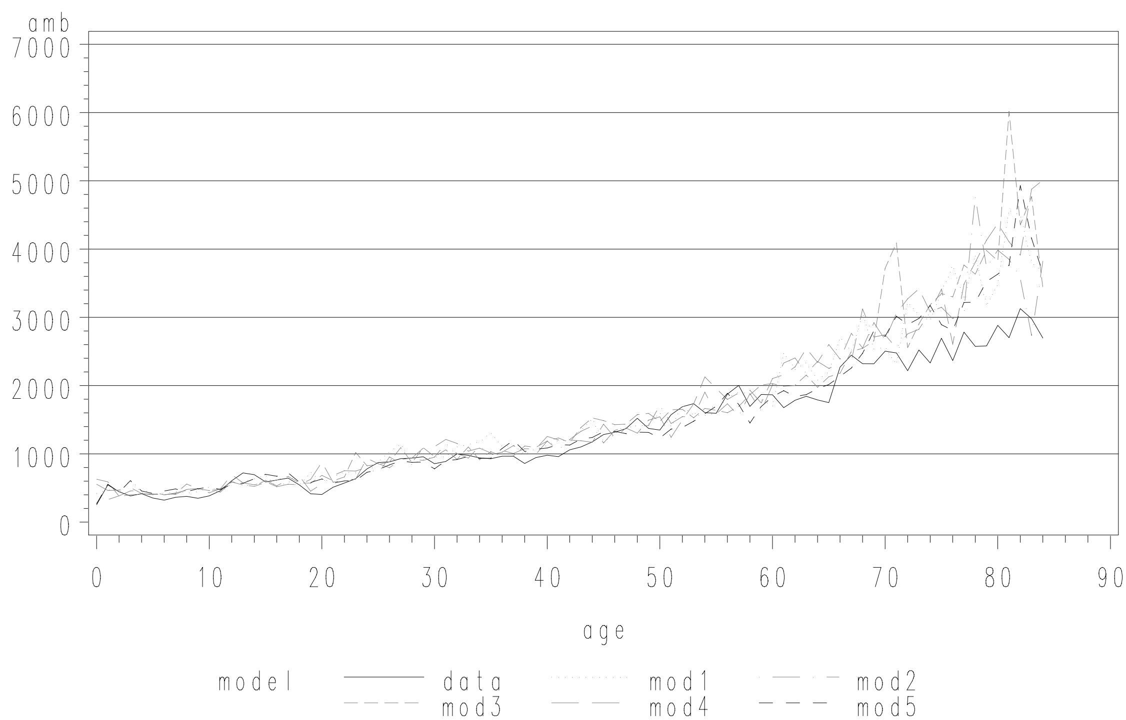


Figure 8: Average of expenditure by age for women (in €)

	Data	Model 1	Model 2	Model 3	Model 4	Model 5
$CORR(D(t), D(t-1))$	*0.41418*	0.04049	0.051891	0.36343	0.44604	0.44070
$CORR(D(t), M(t-1))$	*0.13305*	0.05121	0.035345	0.07210	0.10515	0.13342
$CORR(M(t), D(t-1))$	*0.12622*	0.03627	0.034786	0.07541	0.10409	0.12955
$CORR(M(t), M(t-1))$	*0.68349*	0.11733	0.080699	0.28250	0.49865	0.66174

Table 10: Correlation over the time of ambulatory expenditure, Women

covariates. There exist some constant unobserved factors that have a simultaneous impact on the probability of use and on the amount of expenditure in case of use.

Second, correlations in the time series dimension can be explained by state dependence and individual heterogeneity (among other things like serial correlation in residuals). Like in other fields of micro-economics, to get consistent estimators in a model with state dependence is difficult for theoretical and practical reasons. We suggest to use generalized residual of a simple model in cross-section on the first period to model with parsimony the distribution of individual heterogeneity conditional on initial conditions.

Third, as in the cross-section models if the logarithm of the amount is modeled as a linear function of covariates, we have to take care of the distribution of unobserved heterogeneity. In this paper we maintain the assumption of normality, but we show that in this case the variance of this heterogeneity depends in a large part of the lagged endogenous.

And last, we show that estimations of panel models highly improve the description of the correlation of endogenous in the times series dimension without damaging the distributions in cross-sections.

References

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

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

[3] 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.

[4] Deb P., Trivedi P.K., Zimmer D.M. (2009): Dynamic Cost-offsets of Prescription Drug Expenditures: Panel Data Analysis Using a Copula-based Hurdle Model, *Health, Econometrics and Data Group (HEDG) Working Papers,* WP 09/15.

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

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

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

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

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

[14] Maddala GS. (1985): A survey of the literature on selectivity bias as it pertains to health care markets, In : L.F. Rossiter and R.M. Scheffler eds., *Advances in health economics and health services research*, pp. 3-18.

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

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

[17] 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.

[18] 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.

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

[20] 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.

[21] Seshamani M. and Gray AM. (2004a): A longitudinal study of the effects of age and time to death on hospital costs, *Journal of Health Economics*, 23, pp. 217–35.

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

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

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

[25] 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.

[26] 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.

[27] 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.

6 Appendix :

Table 11: 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
$\rho 1$					0.67**		0,64**		0,64**	
$\rho 2$					-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**	

Note : * means "p-value<10 %", ** means "p-value<5 %"

Table 12: 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**	

Note : * means "p-value<10 %", ** means "p-value<5 %"

G 9001 J. FAYOLLE et M. FLEURBAEY
Accumulation, profitabilité et endettement des entreprises

G 9002 H. ROUSSE
Détection et effets de la multicolinéarité dans les modèles linéaires ordinaires - Un prolongement de la réflexion de BELSLEY, KUH et WELSCH

G 9003 P. RALLE et J. TOUJAS-BERNATE
Indexation des salaires : la rupture de 1983

G 9004 D. GUELLEC et P. RALLE
Compétitivité, croissance et innovation de produit

G 9005 P. RALLE et J. TOUJAS-BERNATE
Les conséquences de la désindexation. Analyse dans une maquette prix-salaires

G 9101 Equipe AMADEUS
Le modèle AMADEUS - Première partie - Présentation générale

G 9102 J.L. BRILLET
Le modèle AMADEUS - Deuxième partie - Propriétés variantielles

G 9103 D. GUELLEC et P. RALLE
Endogenous growth and product innovation

G 9104 H. ROUSSE
Le modèle AMADEUS - Troisième partie - Le commerce extérieur et l'environnement international

G 9105 H. ROUSSE
Effets de demande et d'offre dans les résultats du commerce extérieur manufacturé de la France au cours des deux dernières décennies

G 9106 B. CREPON
Innovation, taille et concentration : causalités et dynamiques

G 9107 B. AMABLE et D. GUELLEC
Un panorama des théories de la croissance endogène

G 9108 M. GLAUDE et M. MOUTARDIER
Une évaluation du coût direct de l'enfant de 1979 à 1989

G 9109 P. RALLE et alii
France - Allemagne : performances économiques comparées

G 9110 J.L. BRILLET
Micro-DMS ***NON PARU***

G 9111 A. MAGNIER
Effets accélérateur et multiplicateur en France depuis 1970 : quelques résultats empiriques

G 9112 B. CREPON et G. DUREAU
Investissement en recherche-développement : analyse de causalités dans un modèle d'accélérateur généralisé

G 9113 J.L. BRILLET, H. ERKEL-ROUSSE, J. TOUJAS-BERNATE
"France-Allemagne Couplées" - Deux économies vues par une maquette macro-économétrique

G 9201 W.J. ADAMS, B. CREPON, D. ENCAOUA
Choix technologiques et stratégies de dissuasion d'entrée

G 9202 J. OLIVEIRA-MARTINS, J. TOUJAS-BERNATE
Macro-economic import functions with imperfect competition - An application to the E.C. Trade

G 9203 I. STAPIC
Les échanges internationaux de services de la France dans le cadre des négociations multilatérales du GATT
Juin 1992 (1ère version)
Novembre 1992 (version finale)

G 9204 P. SEVESTRE
L'économétrie sur données individuelles-temporelles. Une note introductive

G 9205 H. ERKEL-ROUSSE
Le commerce extérieur et l'environnement international dans le modèle AMADEUS (réestimation 1992)

G 9206 N. GREENAN et D. GUELLEC
Coordination within the firm and endogenous growth

G 9207 A. MAGNIER et J. TOUJAS-BERNATE
Technology and trade : empirical evidences for the major five industrialized countries

G 9208 B. CREPON, E. DUGUET, D. ENCAOUA et P. MOHNEN
Cooperative, non cooperative R & D and optimal patent life

G 9209 B. CREPON et E. DUGUET
Research and development, competition and innovation : an application of pseudo maximum likelihood methods to Poisson models with heterogeneity

G 9301 J. TOUJAS-BERNATE
Commerce international et concurrence imparfaite : développements récents et implications pour la politique commerciale

G 9302 Ch. CASES
Durées de chômage et comportements d'offre de travail : une revue de la littérature

G 9303 H. ERKEL-ROUSSE
Union économique et monétaire : le débat économique

G 9304 N. GREENAN - D. GUELLEC / G. BROUSSAUDIER - L. MIOTTI
Innovation organisationnelle, dynamisme technologique et performances des entreprises

G 9305 P. JAILLARD
Le traité de Maastricht : présentation juridique et historique

G 9306 J.L. BRILLET
Micro-DMS : présentation et propriétés

G 9307 J.L. BRILLET
Micro-DMS - variantes : les tableaux

G 9308 S. JACOBZONE
Les grands réseaux publics français dans une perspective européenne

G 9309 L. BLOCH - B. CŒURE
Profitabilité de l'investissement productif et transmission des chocs financiers

G 9310 J. BOURDIEU - B. COLIN-SEDILLOT
Les théories sur la structure optimal du capital : quelques points de repère

G 9311 J. BOURDIEU - B. COLIN-SEDILLOT
Les décisions de financement des entreprises françaises : une évaluation empirique des théories de la structure optimale du capital

G 9312 L. BLOCH - B. CŒURÉ
Q de Tobin marginal et transmission des chocs financiers

G 9313 Equipes Amadeus (INSEE), Banque de France, Métric (DP)
Présentation des propriétés des principaux modèles macroéconomiques du Service Public

G 9314 B. CREPON - E. DUGUET
Research & Development, competition and innovation

G 9315 B. DORMONT
Quelle est l'influence du coût du travail sur l'emploi ?

G 9316 D. BLANCHET - C. BROUSSE
Deux études sur l'âge de la retraite

G 9317 D. BLANCHET
Répartition du travail dans une population hétérogène : deux notes

G 9318 D. EYSSARTIER - N. PONTY
AMADEUS - an annual macro-economic model for the medium and long term

G 9319 G. CETTE - Ph. CUNÉO - D. EYSSARTIER - J. GAUTIÉ
Les effets sur l'emploi d'un abaissement du coût du travail des jeunes

G 9401 D. BLANCHET
Les structures par âge importent-elles ?

G 9402 J. GAUTIÉ
Le chômage des jeunes en France : problème de formation ou phénomène de file d'attente ? Quelques éléments du débat

G 9403 P. QUIRION
Les déchets en France : éléments statistiques et économiques

G 9404 D. LADIRAY - M. GRUN-REHOMME
Lissage par moyennes mobiles - Le problème des extrémités de série

G 9405 V. MAILLARD
Théorie et pratique de la correction des effets de jours ouvrables

G 9406 F. ROSENWALD
La décision d'investir

G 9407 S. JACOBZONE
Les apports de l'économie industrielle pour définir la stratégie économique de l'hôpital public

G 9408 L. BLOCH, J. BOURDIEU, B. COLIN-SEDILLOT, G. LONGUEVILLE
Du défaut de paiement au dépôt de bilan : les banquiers face aux PME en difficulté

G 9409 D. EYSSARTIER, P. MAIRE
Impacts macro-économiques de mesures d'aide au logement - quelques éléments d'évaluation

G 9410 F. ROSENWALD
Suivi conjoncturel de l'investissement

G 9411 C. DEFEUILLEY - Ph. QUIRION
Les déchets d'emballages ménagers : une analyse économique des politiques française et allemande

G 9412 J. BOURDIEU - B. CŒURÉ - B. COLIN-SEDILLOT
Investissement, incertitude et irréversibilité
Quelques développements récents de la théorie de l'investissement

G 9413 B. DORMONT - M. PAUCHET
L'évaluation de l'élasticité emploi-salaire dépend-elle des structures de qualification ?

G 9414 I. KABLA
Le Choix de breveter une invention

G 9501 J. BOURDIEU - B. CŒURÉ - B. SEDILLOT
Irreversible Investment and Uncertainty : When is there a Value of Waiting ?

G 9502 L. BLOCH - B. CŒURÉ
Imperfections du marché du crédit, investissement des entreprises et cycle économique

G 9503 D. GOUX - E. MAURIN
Les transformations de la demande de travail par qualification en France
Une étude sur la période 1970-1993

G 9504 N. GREENAN
Technologie, changement organisationnel, qualifications et emploi : une étude empirique sur l'industrie manufacturière

G 9505 D. GOUX - E. MAURIN
Persistance des hiérarchies sectorielles de salaires: un réexamen sur données françaises

G 9505 Bis D. GOUX - E. MAURIN
Persistence of inter-industry wages differentials: a reexamination on matched worker-firm panel data

G 9506 S. JACOBZONE
Les liens entre RMI et chômage, une mise en perspective
NON PARU - article sorti dans Economie et Prévision n°122 (1996) - pages 95 à 113

G 9507 G. CETTE - S. MAHFOUZ
Le partage primaire du revenu
Constat descriptif sur longue période

G 9601 Banque de France - CEPREMAP - Direction de la Prévision - Erasme - INSEE - OFCE
Structures et propriétés de cinq modèles macro-économiques français

G 9602 Rapport d'activité de la DESE de l'année 1995

G 9603 J. BOURDIEU - A. DRAZNIEKS
L'octroi de crédit aux PME : une analyse à partir d'informations bancaires

G 9604 A. TOPIOL-BENSAÏD
Les implantations japonaises en France

G 9605 P. GENIER - S. JACOBZONE
Comportements de prévention, consommation d'alcool et tabagie : peut-on parler d'une gestion globale du capital santé ?
Une modélisation microéconométrique empirique

G 9606 C. DOZ - F. LENGLART
Factor analysis and unobserved component models: an application to the study of French business surveys

G 9607 N. GREENAN - D. GUELLEC
La théorie coopérative de la firme

G2005/10 P.-O. BEFFY - C. L'ANGEVIN
Chômage et boucle prix-salaires : apport d'un modèle « qualiifés/peu qualifiés »

G2005/11 B. HEITZ
A two-states Markov-switching model of inflation in France and the USA: credible target VS inflation spiral

G2005/12 O. BIAU - H. ERKEL-ROUSSE - N. FERRARI
Réponses individuelles aux enquêtes de conjoncture et prévision macroéconomiques : Exemple de la prévision de la production manufacturière

G2005/13 P. AUBERT - D. BLANCHET - D. BLAU
The labour market after age 50: some elements of a Franco-American comparison

G2005/14 D. BLANCHET - T. DEBRAND - P. DOURGNON - P. POLLET
L'enquête SHARE : présentation et premiers résultats de l'édition française

G2005/15 M. DUÉE
La modélisation des comportements démographiques dans le modèle de microsimulation DESTINIE

G2005/16 H. RAOUI - S. ROUX
Étude de simulation sur la participation versée aux salariés par les entreprises

G2006/01 C. BONNET - S. BUFFETEAU - P. GODEFROY
Disparités de retraite de droit direct entre hommes et femmes : quelles évolutions ?

G2006/02 C. PICART
Les gazelles en France

G2006/03 P. AUBERT - B. CRÉPON -P. ZAMORA
Le rendement apparent de la formation continue dans les entreprises : effets sur la productivité et les salaires

G2006/04 J.-F. OUVRARD - R. RATHELOT
Demographic change and unemployment: what do macroeconometric models predict?

G2006/05 D. BLANCHET - J.-F. OUVRARD
Indicateurs d'engagements implicites des systèmes de retraite : chiffrages, propriétés analytiques et réactions à des chocs démographiques types

G2006/06 G. BIAU - O. BIAU - L. ROUVIERE
Nonparametric Forecasting of the Manufacturing Output Growth with Firm-level Survey Data

G2006/07 C. AFSA - P. GIVORD
Le rôle des conditions de travail dans les absences pour maladie

G2006/08 P. SILLARD - C. L'ANGEVIN - S. SERRAVALLE
Performances comparées à l'exportation de la France et de ses principaux partenaires
Une analyse structurelle sur 12 ans

G2006/09 X. BOUTIN - S. QUANTIN
Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984-2002

G2006/10 C. AFSA
L'estimation d'un coût implicite de la pénibilité du travail chez les travailleurs âgés

G2006/11 C. LELARGE
Les entreprises (industrielles) françaises sont-elles à la frontière technologique ?

G2006/12 O. BIAU - N. FERRARI
Théorie de l'opinion
Faut-il pondérer les réponses individuelles ?

G2006/13 A. KOUBI - S. ROUX
Une réinterprétation de la relation entre productivité et inégalités salariales dans les entreprises

G2006/14 R. RATHELOT - P. SILLARD
The impact of local taxes on plants location decision

G2006/15 L. GONZALEZ - C. PICART
Diversification, recentrage et poids des activités de support dans les groupes (1993-2000)

G2007/01 D. SRAER
Allègements de cotisations patronales et dynamique salariale

G2007/02 V. ALBOUY - L. LEQUIEN
Les rendements non monétaires de l'éducation : le cas de la santé

G2007/03 D. BLANCHET - T. DEBRAND
Aspiration à la retraite, santé et satisfaction au travail : une comparaison européenne

G2007/04 M. BARLET - L. CRUSSON
Quel impact des variations du prix du pétrole sur la croissance française ?

G2007/05 C. PICART
Flux d'emploi et de main-d'œuvre en France : un réexamen

G2007/06 V. ALBOUY - C. TAVAN
Massification et démocratisation de l'enseignement supérieur en France

G2007/07 T. LE BARBANCHON
The Changing response to oil price shocks in France : a DSGE type approach

G2007/08 T. CHANEY - D. SRAER - D. THESMAR
Collateral Value and Corporate Investment Evidence from the French Real Estate Market

G2007/09 J. BOISSINOT
Consumption over the Life Cycle: Facts for France

G2007/10 C. AFSA
Interpréter les variables de satisfaction : l'exemple de la durée du travail

G2007/11 R. RATHELOT - P. SILLARD
Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?

G2007/12 V. ALBOUY - B. CRÉPON
Aléa moral en santé : une évaluation dans le cadre du modèle causal de Rubin

G2008/01 C. PICART
Les PME françaises : rentables mais peu dynamiques

G2008/02 P. BISCOURP - X. BOUTIN - T. VERGÉ
The Effects of Retail Regulations on Prices Evidence form the Loi Galland

G2008/03 Y. BARBESOL - A. BRIANT
Économies d'agglomération et productivité des entreprises : estimation sur données individuelles françaises

G2008/04 D. BLANCHET - F. LE GALLO
Les projections démographiques : principaux mécanismes et retour sur l'expérience française

G2008/05 D. BLANCHET - F. TOUTLEMONDE
Évolutions démographiques et déformation du cycle de vie active : quelles relations ?

G2008/06 M. BARLET - D. BLANCHET - L. CRUSSON
Internationalisation et flux d'emplois : que dit une approche comptable ?

G2008/07 C. LELARGE - D. SRAER - D. THESMAR
Entrepreurship and Credit Constraints - Evidence from a French Loan Guarantee Program

G2008/08 X. BOUTIN - L. JANIN
Are Prices Really Affected by Mergers?

G2008/09 M. BARLET - A. BRIANT - L. CRUSSON
Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu

G2008/10 M. BEFFY - É. COUDIN - R. RATHELOT
Who is confronted to insecure labor market histories? Some evidence based on the French labor market transition

G2008/11 M. ROGER - E. WALRAET
Social Security and Well-Being of the Elderly: the Case of France

G2008/12 C. AFSA
Analyser les composantes du bien-être et de son évolution
Une approche empirique sur données individuelles

G2008/13 M. BARLET - D. BLANCHET - T. LE BARBANCHON
Microsimuler le marché du travail : un prototype

G2009/01 P.-A. PIONNIER
Le partage de la valeur ajoutée en France, 1949-2007

G2009/02 Laurent CLAVEL - Christelle MINODIER
A Monthly Indicator of the French Business Climate

G2009/03 H. ERKEL-ROUSSE - C. MINODIER
Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth?
A Real-Time Analysis on French Data

G2009/04 P. GIVORD - L. WILNER
Les contrats temporaires : trappe ou marchepied vers l'emploi stable ?

G2009/05 G. LALANNE - P.-A. PIONNIER - O. SIMON
Le partage des fruits de la croissance de 1950 à 2008 : une approche par les comptes de surplus

G2009/06 L. DAVEZIES - X. D'HAULTFOEUILLE
Faut-il pondérer ?… Ou l'éternelle question de l'économètre confronté à des données d'enquête

G2009/07 S. QUANTIN - S. RASPILLER - S. SERRAVALLE
Commerce intragroupe, fiscalité et prix de transferts : une analyse sur données françaises

G2009/08 M. CLERC - V. MARCUS
Élasticités-prix des consommations énergétiques des ménages

G2009/09 G. LALANNE - E. POULIQUEN - O. SIMON
Prix du pétrole et croisssance potentielle à long terme

G2009/10 D. BLANCHET - J. LE CACHEUX - V. MARCUS
Adjusted net savings and other approaches to sustaibability: some therotical backgroud

G2009/11 V. BELLAMY - G. CONSALES - M. FESSEAU - S. LE LAIDIER - É. RAYNAUD
Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003

G2009/12 J. BARDAJI - F. TALLET
Detecting Economic Regimes in France : a Qualitative Markov-Switching Indicator Using Mixed Frequency Data

G2009/13 R. AEBERHARDT - D. FOUGÈRE - R. RATHELOT
Discrimination à l'embauche : comment exploiter les procédures de *testing* ?

G2009/14 Y. BARBESOL - P. GIVORD - S. QUANTIN
Partage de la valeur ajoutée, approche par données microéconomiques

G2009/15 I. BUONO - G. LALANNE
The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade

G2010/01 C. MINODIER
Avantages comparés des séries des premières valeurs publiées et des séries des valeurs révisées - Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France

G2010/02 V. ALBOUY - L. DAVEZIES - T. DEBRAND
Health Expenditure Models: a Comparison of Five Specifications using Panel Data