HEDG Health, Econometrics and Data Group
THE UNIVERSITY of York

HEDG Working Paper 09/03

The Effect Of Supplemental Insurance On Health Care Demand With Multiple Information: A Latent Class Analysis

VALENTINO DARDANONI
PAOLO LI DONNI

March 2009

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

THE EFFECT OF SUPPLEMENTAL INSURANCE ON HEALTH CARE DEMAND WITH MULTIPLE INFORMATION: A LATENT CLASS ANALYSIS

VALENTINO DARDANONI AND PAOLO LI DONNI

Abstract. The Medicare program, which provides insurance coverage to the elderly in the United States, does not protect them fully against high out-of-pocket costs. For this reason private supplementary insurance, named Medigap, has been available to cover Medicare gaps. This paper studies how Medigap affects the utilization of health care services. The decision to take out supplemental insurance is likely to be influenced by unobservable attributes such as actual risk type and insurance preferences. Empirical appraisals to this problem typically rely on the recursive bivariate probit. We exploit the Health and Retirement Study data and some recent advances on latent class analysis to jointly model the insurance and health care decisions. Results show the presence of unobserved 'types' representing different preferences and risk levels. We compare our results to those obtained by the probit and the bivariate probit and find the residual effect of insurance on health care not significant.

JEL Classification Numbers C52, D82, G22, I10.

Keywords Health Care Demand; Latent Class Models; Health Insurance; Asymmetric Information; Medigap.

1. Introduction

Medicare is a public program which provides health insurance for the elderly (aged 65 or older) and some disabled non elderly. As many other standard health insurance plans, Medicare relies deeply on mechanisms such as coinsurance, deductibles and copayments to control health care expenditure for many covered services. This insurance structure leaves beneficiaries at risk for large out-of-pocket expenses. As a result, many beneficiaries purchase voluntary supplemental private policies, such as Medigap, to fill Medicare's gaps in non-covered health care services and limit cost sharing.

Medicare cost-sharing structure reflects the belief that health insurance, by lowering the price per services, gives individuals' an *incentive* to increase the demand for health care. Although its presence - usually called *ex-post moral hazard* - is very well known by the theoretical literature on contract theory (Arrow[1], Pauly [36] and Zweifel and Manning [43]), it is still debated empirically because of the existence of *self-selection*, since individuals

Date: March 5, 2009.

who expect high health care costs may choose a more generous coverage and then ex-post purchase more services. This means that individual choice to take out a voluntary health insurance contract is driven by unobserved attributes which potentially affect health care utilization. Empirically it is a serious problem to distinguish between the incentive and the selection effect, since they would lead typically the same kind of observed (positive) correlation between insurance coverage and risk occurrence.

There are different ways to distinguish empirically the selection from the incentive effect. A strategy is to use experimental data such as the RAND Health Insurance Experiment (RHIE), where to identify the incentive effect controlling for the self-selection individuals were randomly assigned to plans with different coverages so that insurance choice becomes exogenous. Another strategy is to exploit (quasi) natural experiments where insurance choice or the incentive structure has been modified exogenously (Chiappori *et al.* [8] and Eichner [18]).

In observational studies traditional empirical analyses of health care consumption tend to consider insurance choice as exogenous and independent from the decision of medical care use (Ettner [19]). This approach relies on the belief that observables in the regression fully capture all sources that can potentially affect insurance choice. In particular a standard approach to deal with the self-selection issue is to model endogenously the insurance choice and estimate a simultaneous two equations bivariate probit model with a recursive structure between insurance and health care utilization (Holly, *et al.* [26], Jones *et al.* [29], Bago d'Uva and Santos Silva [2]). This estimation strategy may lead to puzzling results - such as *advantageous selection* - since risk occurrence may be negatively related to unobserved factors (e.g. actual health risk, preferences for risk and attitudes on health care utilization and insurance purchase) which positively influence insurance choice.

Recent contribution to this area point to this problem as the existence of multiple dimension of private information (Finkelstein and McGarry [22]). To account for this issue, a more convincing distinction may be between individual "types", the difference being determined by the unobserved factors mentioned above. This framework is particularly suitable for latent class analysis (LCA) which has been exploited to analyse the health care demand (Deb and Trivedi [15]-[16]). This estimation strategy provides a convincing alternative to the bivariate probit in the analysis of insurance market since it can identify different groups of individual ("types") sharing the same unobserved characteristics in terms of risk attitude and actual health risk.

This paper is an attempt to measure the *effect* of supplemental insurance on health care demand - namely the incentive effect - by conditioning on individual unobserved "types". Such conditioning is made possible by exploiting some recent developments on finite mixture models (Huang and Bandeen-Roche [27], Bartolucci and Forcina [3], Dardanoni *et al.* [14]), which allow response variables to depend on covariates and residual association between responses.

To analyse the "incentive" effect we consider as dependent variable a binary variable representing any hospital admission in the previous two years which account for 29% of the Medicare's total expenditure.[1] At this purpose we use the Health and Retirement Survey (HRS) dataset to control for individual risk preferences and actual risk. This is achieved by exploiting the panel features of HRS, which contains information on a rich set of variables concerning health status and individual preferences for risk. The empirical finding we report confirm the presence of multiple sources of private information affecting the insurance choice and the inpatient hospital admission. However, we find no evidence of direct effect of Medigap.

The paper is organized as follows. In the next section we report a brief overview of Medicare and Medigap insurance contracts; section 3 reviews the main empirical contribution in the related literature; we then discuss the model to be estimated (section 4) using the data described in section 5. Finally section 6 and 7 report the main findings and some concluding remarks respectively. In the Appendix A we explain the details of the identification and estimation of the model.

2. Health insurance and access to care for elderly in US

2.1. **Medicare.** Medicare is probably the main source of health insurance for all individuals aged 65 and the coverage is near universal (about 97% of the elderly have Medicare)[2].

Medicare programme consists mainly of two plans in which people may be enrolled. The first plan, named Medicare Part A, is also known as "Hospital Insurance" since it covers the basic hospital's health care services such as inpatient's admissions. Most of beneficiaries, who have paid Medicare taxes for at least 10 years, are automatically enrolled with their spouse in Part A when they turn 65. Part A plan pays almost the entire medical expenditure (except a deductible) for the first 60 nights of inpatient hospital staying and imposes an increasing cost sharing structure if hospital admission lasts over this first period.

The second plan is Medicare Part B. Most of beneficiaries choose to extend Medicare Part A insurance coverage to Part B because it covers several medicare services such as doctors' services, outpatient care and some preventive services. Part B enrollment requires the payment of a monthly premium which may depend on income. Part B's deductible and co-payment amount respectively to $110 and the 20 percent of expense in 2005 exceeding the deductible for all Medicare approved services.

[1]CMS, Office of the Actuary, National Health Expenditure Accounts, 2007.

[2]Current Population Reports (2005) "Income, Poverty and Health Insurance Coverage in the United States: 2004"

2.2. **Supplemental insurance coverage and medigap policy.** There are several limitations of Medicare original plans: limitation in the coverage of health care services, high out-of-pocket expenses to beneficiaries and lack of a catastrophic cap expenditure. These induce seniors to seek additional coverage provided by private insurance.

There are three main sources of supplemental private insurance which pay for some additional (to Medicare) services or help pay the share of the costs of Medicare-covered services. The first one is the employer-sponsored supplemental insurance and it is purchased usually by a former employer or union. The second one is represented by Tricare (available only to military personal) and the Medicare Advantage plans (Part C) provided by by private health insurance.

The third one and also the most common source of supplemental coverage comes from Medigap-private health insurance which are specifically designed to cover those "gaps" of coverage left by original Medicare plans. Since 1990 Medigap insurance market is highly regulated by the Federal law. Medigap plans are standardized into ten plans,"A" through "J", which cover a single individual, offer certain additional services and help beneficiaries pay health care cost (deductibles and co-payment) that the original Medicare plan does not cover. This means that if individuals are enrolled in Medicare plus a supplemental Medigap insurance health care cost is covered by both plans. For example the basic plan, A, covers the entire coinsurance or copayments for hospital stays, physician visits and outpatient care.

Federal regulation of Medigap market designed a particular mechanism favoring the insured: Medigap insurance companies must offer also the basic plan "A" if they offer any other more generous plan. In addition, there is a free enrolment period which lasts for six months from the first month in which people are both 65 years old and enrolled in Medicare Part B. During this period Medigap cannot refuse any insurer even if there are pre-existing conditions. Legal restrictions involve also the pricing criteria, which are mainly based on individual's age.

3. Related Literature

The empirical literature on the incentive and selection effect in health insurance is continuously growing and controversial since disentangling the two effects is not straightforward because the unobserved nature of individual preferences and health status pose serious endogeneity problems.

A "radical" solution is to exploit experiments or some particular features of the data which render insurance choice exogenous. The best known study is the RAND Health Insurance Experiment conducted in the 1974. To control for self-selection individuals were randomly assigned to insurance plans with different coinsurance rate. Manning *et. al.* [35] show that patients ensured by a plan with first dollar coverage had 37% more physician visits than those facing co-insurance rates of 25% suggesting strong evidence of *ex-post* moral hazard, but they found no significant incentive effect on inpatient care.

In non-experimental settings most of the studies use large observational data sets which include information on individuals, health care services and insurance status. There are different econometric strategies to empirically appraise this issue. The first approach considers insurance choice as exogenous in the health care utilization equation, and estimates health care utilization or medical expenditure with a probit model (Hurd and McGarry [28]) or a two-parts model (Ettner [19], Khandker and McCormack [30]-[31]) using health indicators to mitigate the presence of unobserved heterogeneity in health status. In general results show that Medicare enrollees with supplemental insurance (Medigap or employer plans) show the highest levels of spending and that individuals reporting better health were significantly more likely to enroll in private supplemental plans. Another approach is to model endogenously insurance choice considering both selection on observable and unobservable factors. In this framework many studies conducted in the European health insurance market exploited a recursive bivariate probit to model simultaneously the probability to have at least one inpatient stay and purchasing supplemental insurance (Holly *et al.* [26], Jones *et al.* [29], Buchmueller *et al.* [5]). In general the most common finding in these empirical studies exploiting the bivariate probit model is to find a positive (direct) effect of insurance on health care demand and no positive (statistically significant) correlation between residuals of the insurance and the health risk occurrence equations respectively.

These findings are arousing a great deal of interest among researchers. In particular Finkelstein and McGarry [22] and Cutler *et al.* [11] take an innovative approach based on insurance company unused variables to test the (positive) correlation between health care utilization and insurance coverage. Using one wave of the Health Dynamics Among Oldest (AHEAD) Finkelstein and McGarry [22] identify two groups of individuals who purchase long-term insurance and use nursing home: those who prefer insurance for cautionary reasons and ex-post are less likely to enter a nursing home and those who are subjectively riskier and ex post have higher risk occurrence.[3] They find a (no significant) negative correlation and suggest that this result may be related to multiple dimension of private information, since advantageous selection based on risk preference and adverse selection based on actual risk offset each other.[4] Cutler *et al.* [11] confirm these findings in the Medigap insurance market; Fang *et al.* [20] provide strong evidence of advantageous selection in this market and find cognitive ability is an important factor influencing selection. They conclude that this result reflect the idea that senior citizens may have difficulties in understanding Medicare and Medigap rules.

[3]AHEAD is a cohort of the Health and Retirement Study (HRS) from which our sample is drawn.

[4]Evidence of positive and statistical significant relationship between risk aversion and health attitudes have been found in the US health insurance market by Vistnes and Banthin [42], Landerman *et al.* [33]

A rather different approach is to control for unobserved heterogeneity using LC analysis and keeping insurance choice exogenous. Deb and Trivedi [15]-[16] develop a finite mixture negative binomial and estimate health care demand for several health care measures. They distinguish two unobserved groups of population: the "healthy" and the "ill". After controlling for these two unobserved "types" of people, they find that individuals with supplementary private health insurance tend to seek care from physicians and non-physicians more often than the uninsured, while this effect is not significant for hospital utilization.

In this paper we use recent advances in LC analysis to control for unobserved heterogeneity and estimate the insurance effect on hospital utilization. Although Finkelstein and McGarry [22] focus mainly on asymmetric information rather than the incentive effect, what makes their paper similar to ours is the role played by the multiple dimension of individual's private information in the insurance choice.

4. The model

Let M_t denote a binary variable which takes value 1 if an individual uses some medical care at time t, and S_t a binary variable which takes value 1 if an individual has supplementary insurance at time t. We want to study the following conditional expectations:

$$Pr(M_t = 1 \mid S_t, \boldsymbol{w}_M), \tag{1}$$

$$Pr(S_t = 1 \mid \boldsymbol{w}_S) \tag{2}$$

where $\boldsymbol{w}_M$ and $\boldsymbol{w}_S$ are vectors of individual characteristics which affect respectively medical care use and supplementary insurance purchase; $\boldsymbol{w}_M$ and $\boldsymbol{w}_S$ include the determinants of individual's medical care and insurance choices, and in particular individuals' preferences and constraints. Thus, $\boldsymbol{w}_M$ and $\boldsymbol{w}_S$ are likely to include both observable individuals' characteristics and a set of unobservable characteristics such as risk tolerance, attitude towards medical care use and insurance purchase, actual (health) riskiness and so on. If one could control properly for $\boldsymbol{w}_M$, than the effect of supplementary insurance on medical care use can be directly obtained by an appropriate binary regression model. However, since unobservables enter both (1) and (2), estimation of the effect of S_t on M_t is fraught with the endogeneity problem.

To capture unobservable heterogeneity, a successful strategy in applied research is to use lagged dependent variables; in this application, lagged values of M_t and S_t may act as proxies for unobservable attitudes to buy insurance and use health care. Letting $\boldsymbol{x}_M$ and $\boldsymbol{x}_S$ denote vectors of observable characteristics, the following system may be considered as a first step towards modeling (1,2):

$$M_t^* = \alpha^{M_t} + \beta^{M_t} S_t + \gamma^{M_t} M_{t-1} + \boldsymbol{x}'_M \boldsymbol{\delta}^{M_t} + U + \epsilon_M \tag{3}$$

$$S_t^* = \alpha^{S_t} + \beta^{S_t} S_{t-1} + \gamma^{S_t} M_{t-1} + \boldsymbol{x}'_S \boldsymbol{\delta}^{S_t} + U + \epsilon_S \tag{4}$$

where as usual M_t^* and S_t^* denote the unobservable continuous counterparts of $M_t = 1(M_t^*)$ and $S_t = 1(S_t^*)$, U denotes any residual unobservable heterogeneity, and ϵ_M and ϵ_S idiosyncratic errors. If we now let $\eta_S = U + \epsilon_S$ and $\eta_M = U + \epsilon_M$ and assume that (η_M, η_S) are distributed as a bivariate normal with standard margins and correlation coefficient equal to ρ, we get to the the recursive bivariate probit which can be considered as the workhorse in this literature (see e.g. Jones *et al.* [29]):

$$M_t^* = \alpha^{M_t} + \beta^{M_t} S_t + \gamma^{M_t} M_{t-1} + \boldsymbol{x}'_M \boldsymbol{\delta}^{M_t} + \eta_M \tag{5}$$

$$S_t^* = \alpha^{S_t} + \beta^{S_t} S_{t-1} + \gamma^{S_t} M_{t-1} + \boldsymbol{x}'_S \boldsymbol{\delta}^{S_t} + \eta_S \tag{6}$$

The recursive bivariate probit model (5,6) is simple to estimate and to interpret, allows estimation of the effect of supplementary insurance on medical care use with standard software, and has been much used in this context ([5],[26], [29]). However it relies on bivariate normality to achieve parameters' identification, and does not disentangle the different sources of the multiple dimension of the unobserved U (Finkelstein and McGarry [22]) .

Our strategy is to try to control for the residual unobserved heterogeneity U by identifying a finite number of unobservable "types" which differ with respect to their attitude to buy insurance and to use medical care. In particular, we assume that U is a discrete random variable taking values in, say, $\{1, \ldots, m\}$, which define m unobservable heterogeneous "types"; in practice, such as U can be seen as a cross-classification of underlying unobservable individual characteristics, such as risk tolerance and needs and attitudes to use medical care. Equations (3,4) can be written as

$$M_t^* = \sum_{u=1}^{m} \alpha_u^{M_t} U_u + \beta^{M_t} S_t + \gamma^{M_t} M_{t-1} + \boldsymbol{x}'_M \boldsymbol{\delta}^{M_t} + \epsilon_M \tag{7}$$

$$S_t^* = \sum_{u=1}^{m} \alpha_u^{S_t} U_u + \beta^{S_t} S_{t-1} + \gamma^{S_t} M_{t-1} + +\boldsymbol{x}'_S \boldsymbol{\delta}^{S_t} + \epsilon_S \tag{8}$$

where $U_1, \ldots, U_m$ denote the set of m dummy variables indicating "latent type" membership. Thus, the coefficients $\alpha_u^{M_t}$ and $\alpha_u^{S_t}$ can be interpreted as random intercepts with a nonparametric discrete specification, like in Heckman and Singer [25].

In order to identify U, we exploit the dynamic structure of individuals' choices, and in particular lagged values of M and S are used not only *directly* in equations (7,8), but also *indirectly* as *indicators* of U. Moreover, we also use three other binary indicators, namely health status (H_1), smoking behavior (H_2) and gender-appropriate preventive health care (H_3), defined below in section 5, which are meant to be further indicators of unobservable individual characteristics such as risk tolerance, health needs and attitudes to use medical care.

Using a standard logit link in equations (7-8), we estimate the model:

$$\begin{aligned} \lambda^{M_t} &= \textstyle\sum_{u=1}^{m} \alpha_u^{M_t} U_u + \beta^{M_t} S_t + \gamma^{M_t} M_{t-1} + \boldsymbol{x}'_M \boldsymbol{\delta}^{M_t} \\ \lambda^{S_t} &= \textstyle\sum_{u=1}^{m} \alpha_u^{S_t} U_u + \beta^{S_t} S_{t-1} + \beta^{S_t} M_{t-1} + \boldsymbol{x}'_S \boldsymbol{\delta}^{S_t} \end{aligned} \tag{9}$$

together with the class membership probabilities $Pr(U = u)$ which can be written in terms of *adjacent logits* as

$$\log\left(\frac{Pr(U=u+1)}{Pr(U=u)}\right) = \lambda_u^U = \alpha_u^U \qquad u = 1, \ldots, m-1 \tag{10}$$

and in addition to (9-10), the following recursive system which can be considered instrumental for identifying U:

$$\begin{aligned}
\lambda^{M_{t-1}} &= \textstyle\sum_{u=1}^{m} \alpha_u^{M_{t-1}} U_u + \beta^{M_{t-1}} S_{t-1} + \gamma^{M_{t-1}} M_{t-2} \\
\lambda^{S_{t-1}} &= \textstyle\sum_{u=1}^{m} \alpha_u^{S_{t-1}} U_u + \beta^{S_{t-1}} S_{t-2} + \gamma^{S_{t-1}} M_{t-2} \\
\lambda^{M_{t-2}} &= \textstyle\sum_{u=1}^{m} \alpha_u^{M_{t-2}} U_u + \beta^{M_{t-2}} S_{t-2} \\
\lambda^{S_{t-2}} &= \textstyle\sum_{u=1}^{m} \alpha_u^{S_{t-2}} U_u \\
\lambda^{H_j} &= \textstyle\sum_{u=1}^{m} \alpha_u^{H_j} U_u \qquad j = 1, 2, 3
\end{aligned} \tag{11}$$

Notice that lagged values of insurance choice and hospital utilization serve a double duty in equations (9-11): they are used directly in equation (9) to capture persistency, and in the system of auxiliary equations (11) for the purpose of identification of the unobservable types. Letting now $\boldsymbol{\lambda}$ collect the set of logits, equations (9-11) can be compactly written as

$$\boldsymbol{\lambda} = \boldsymbol{X}\boldsymbol{\psi} \tag{12}$$

where $\boldsymbol{\psi}$ collects the model parameters α's, β's, γ's and δ's. Estimation of the model parameters $\boldsymbol{\psi}$ can be implemented by the EM algorithm; Appendix A discusses estimation and identification of the model.[5]

An interesting feature of the estimation algorithm is that equality and inequality constraints can be imposed in the estimation of the model parameters. Thus, since standard contract theory (see e.g. Finkelstein [21]) predicts that in the Medigap insurance market with a pooling equilibrium ill-health individuals tend both to purchase more insurance coverage and use ex-post more health care, if one takes as a point of departure the standard insurance model where individuals have private information on their risk types and so U is a unidimensional unobservable variable which captures individual's actual riskiness, then rearranging the m types according to their actual health one should observe that *both* insurance choice *and* health care utilization will be decreasing in U. Thus, a test of the unidimensionality of private information can be performed as follows:

- order the m types according to the probability of having a good health status H_1, that is, order U so that $\alpha_1^{H_1} \geq \cdots \geq \alpha_m^{H_1}$ (note that finite mixture models are invariant to the rearrangement of the types);
- test whether hospital use M_t and insurance choice S_t are decreasing in U, that is, test whether $\alpha_1^{M_t} \leq \cdots \leq \alpha_m^{M_t}$ and $\alpha_1^{S_t} \leq \cdots \leq \alpha_m^{S_t}$;

[5] We are grateful to Antonio Forcina for kindly providing the Matlab code for the estimation.

- the null hypothesis that there is a unidimensional underlying unobservable variable U representing actual risk types can thus be tested by setting a system of linear inequalities.

Techniques of order restricted inference can be used to show that the likelihood ratio test statistic for the monotonicity null is asymptotically distributed as a mixture of chi-squared distributions (see Gourieroux and Monfort [24] for a general exposition, Dardanoni and Forcina [12] for an explanation of how the mixing weights can be calculated by simulations, and Kodde and Palm [32] for bounds on the test distribution). Rejection of the null hypothesis implies the presence of multiple dimensions of unobserved heterogeneity, and the existence of individual "types" with differences being driven not only by actual health risk but also by their preferences.

5. Descriptive Statistics

We use data from the Health and Retirement Study (HRS). Since 1992 the HRS is a biennial survey targeting elderly Americans over the age of 50 sponsored by the National Institute on Aging. Although the survey is not conducted on an yearly basis, from 1998 it provides longitudinal data for an array of information, consistently administrated, on several different fields such as health and health care utilization, type of insurance coverage, socioeconomic condition, retirement plans and family structure and transfers. The longitudinal feature of the data offers a suitable dynamic framework to test asymmetric information in health insurance market considering the intertemporal dependence between insurance choice and health care utilization (Chiappori [9]). For our purpose we use the last available wave on 2006 ($t = 3$) as reference point to collect information on insurance status, health and health care utilization from the previous two waves (2002 and 2004). To evaluate more closely the effect of asymmetric information on Medicare expenditure, we consider a sample restricted to Medicare Part A or B enrollees over the last wave. This means that we consider only individuals older than 65 in 2002. Information about Medicare is binary coded and it is clearly reported in the survey as the first question asked in the insurance section.

Since we study the effect of supplemental insurance (Medigap) on health care, we also exclude those individuals that received additional coverage through a former employer, spouse or some other government agency. Following Fang *et al.* [20] we define an individual as having additional health insurance coverage (Medigap) if s/he purchased directly health insurance policy in addition to Medicare. As result we limit the sample to people who deliberately choose to have a supplemental coverage and pay the required monthly premium. Additional piece of information on individual insurance status is available in the survey. In particular the HRS asked respondents whether they were covered by Medicaid, CHAMPUS or CHAMPVA (Tri-care). Thus we have the chance to control also for these sources of insurance.

Our sample is composed by 2931 individuals and descriptive statistics are reported in table 1. We follow previous studies on health care demand (see e.g. Cameron *et al.* [7], Jones *et al.* [29], Deb and Trivedi [16] Vera-Heranandez [41]) and insurance choice (see e.g., Propper [37]-[38], Cameron and Trivedi [6]) as guidance in selecting five groups of variables describing individual socioeconomic characteristics, insurance status, health care consumption, health status and individual risk preferences.

Supplemental insurance status at 2002 (spins02), 2004 (spins04) and 2006 (spins06) is coded as binary variable which takes 1 if respondent has any (no long-term care) supplemental private Medigap insurance coverage. Almost 50 percent of Medicare beneficiaries in the sample has a supplemental insurance and 87 percent of them has this coverage since when they turn 65 (year 2002). Therefore most of medicare beneficiaries purchase a supplemental insurance coverage as soon as they are enrolled in Medicare Part A or B. This is also supported by the fact that insurance status does not vary deeply across waves. Comparing the sub-sample averages for the two groups of people with and without supplemental insurance we find also that beneficiaries of additional insurance have higher education and they are also in the top wealth quartile. In addition to these variables we also use information on whether additional coverage was provided by a former employer (iemp04) or by the spouse (iemps04). These variables have been used in the literature to explain individual choice to take out voluntary supplemental coverage (Jones *et al.* [29]).

The HRS offers detailed information on health care consumption. In particular we focus on hospital staying over the three waves (h02, h04 and h06). These variables are binary and take value 1 if individual had at least one hospital admission, and 0 otherwise. Health conditions have important influence both on the decision to subscribe supplementary insurance as well as on the utilization of health care services. In HRS health condition is measured along different measures. We include in the analysis either self-reported longstanding or chronical disease - such as high blood pressure, hypertension, cardiovascular disease, lung disease, kidney conditions, emotional and psychiatric problems. In addition to this self-reported indicator we also consider mobility limitation based on several aspect of physical health. In particular we construct an indicator which takes 1 if person does not report any functional limitation based on the indexes of Activities of Daily Living (ADL), the Instrumental Activities of Daily Living (IADL), the large muscle index, a mobility index, the fine motor index and the gross motor index uses the walking one block, walking across a room, climbing one flight of stairs, getting in or out of bed, and bathing activities.[6] For our purpose we collapse these indexes

[6]These indexes are developed by the RAND Corporation from HRS public data releases; details can be found in [39]

in a binary indicator (he06) which takes 1 if individual report more than 2 limitations or chronic diseases.

Finally the last group of variables is devoted to capture, jointly with past insurance status, individual precautionary behavior. We follow Finkelstein and McGarry [22] and Cutler *et al.* [11] to measure individual's cautiousness by her investment in risk-reducing activities. In particular we use a binary variable which considers whether the individual is currently a smoker (smoken06), and a binary variable (prev) which captures gender-appropriate preventive health care. These preventive care activities are: whether the individual had a flu shot, had a blood test for cholesterol, checked her breasts for lumps monthly, had a mammogram or breast x-ray, had a Pap smear, and had a prostate screen. The median individual undertakes 60 percent of gender relevant activities; 5 percent report doing nothing and 38 percent report engaging in all relevant activities. For our purpose the preventive behavior is coded as one if the individual undertook at least the 50% of the total amount of gender relevant measures.[7]

6. Results

In this section we present the results for the effect of insurance on utilization. We compare a simple binary probit model for the probability of having a hospital admissions, which does not allow for the endogeneity of insurance, with estimates that do allow for endogeneity under varying assumptions: the bivariate probit model and the extended LCA.

6.1. **Probit and bivariate probit results.** The first two columns of table 2 show the probit model estimates of insurance choice and inpatient hospital admission. Turning first to the controls, the health status indicator has a statistically significant effect on utilization but not on insurance choice. This result is expected given that Medigap monthly premiums are issue-age-rated and then only the age affects how much individuals pay (Cutler *et al.* [11]). Utilization is significantly lower for women than male and it increases with individual age. The effect of wealth on insurance is positive and statistically significant while it has no substantial effect on health care utilization. Moreover individuals with other sources of (public) private insurance have lower probability to buy additional coverage. This effect is more significant if compared with the role played by total wealth. In fact those with lower health may also been covered by public social plan - such as Medicaid - which can offer additional coverage and then in practice they work as substitute of Medigap plans (Finkelstein [21]).

The effect of insurance on hospital admission is large and statistically significant. The results in column 1 show that after controlling for observables individual characteristics,

[7]Following Fang *et al.* [20], we tried to use as indicators also the set of cognitive abilities variables used by them. However, doing so produced a dramatic drop of more than 75% of our sample. Thus, cognitive ability is only implicitly addressed by our use of past history to identify unobservable types.

supplemental insurance increases the probability of having at least one hospital staying. Following Chiappori and Salanie [10] we test the null hypothesis of residual correlation being zero between the insurance and health care utilization probit models. The test statistic is equal to 27.2 which is asymptotically distributed as a $\chi^2(1)$. Thus we can strongly reject the conditional independence hypotheses and then hospital utilization and supplemental insurance are not independent in this framework.

For this reason we then estimate a bivariate probit model with a recursive structure in order to obtain the effect of supplemental insurance coverage on health care utilization. As reported in table 2, the bivariate probit confirms that the effect of insurance is positive (0.94) and statistically significant at 5 per cent even when insurance choice is endogenously determined. Turning now to the controls, we obtain the same result as in the probit model for health status and demographic characteristics. Controlling for risk preference observable proxies support the idea that individuals with more cautious behavior tend to purchase additional insurance coverage and have lower probability to have an hospital admission (Cutler *et al.* [11]). In fact all else equal, individuals with a stable additional coverage in the past years as well as coverage from other (non-public) sources are also those who buy more coverage but that ex post tend to use less resources. On the contrary people covered by public insurance plan buy less additional coverage and tend use more health care. For the risk reducing behavior results are mixed. In fact those who are current smokers without a gender-appropriate preventive health care investment in the previous years use more resources but have lower probability to buy health insurance even if this effect is not statistically significant.

Finally, our estimate of the correlation between residuals (ρ) is negative and equal to -0.43. The estimated standard-error is 0.18 so that we can reject the conditional independency hypothesis concluding that conditional on observables there is a strong self-selection - i.e. advantageous selection (Buchmueller *et al.* [5], Fang *et al.* [20], Jones *et al.* [29]) - in the insurance choice which cannot be neglected. This result is also supported by estimating the restricted model with $\rho = 0$. The test is equal to 4.34 and distributed as a $\chi^2(1)$. Also in this case the restriction of correlation being zero is rejected. Since this result might be related to the presence of multiple dimension of private information we explore this hypothesis by estimating the system of equations (12) which measure the "incentive" effect of insurance on hospital staying directly controlling for selection on unobservables.

6.2. **Results from model** (12)**.** We start by estimating model (12) under different numbers m of latent classes. Maximum likelihood estimation is performed by a EM algorithm as described in the Appendix. Results for $m = 2, 3, 4, 5$ are reported in tables 4-8. Table 3 reports the maximized log-likelihood $L(\psi)$, the Schwartz's Bayesian Information Criterion $BIC(\psi) = -2L(\psi) + \upsilon log(n)$ and the Akaike's information criterion $AIC(\psi) = 2\upsilon - 2log(L(\psi))$, where n denotes sample size and υ is the number of parameters.

BIC and AIC information criteria seem to indicate that three or four LC are adequate to represent the unobserved heterogeneity U. A glance at all tables reveals that estimated β and γ coefficients do not seem to vary substantially with respect to the number of LC specifications.

Let us consider first table 7, which shows that after controlling for unobserved heterogeneity the effect of additional insurance on hospital utilization - namely the incentive effect (β^{M_3}) - does not vary much in magnitude and it is not statistical significant for each LC specification. This evidence contrasts with the probit and bivariate probit estimates suggesting that the traditional empirical models relying on selection on observables may not fully capture the underlying unobserved heterogeneity driven by individual's preferences and actual risk. Moreover the residual effect of supplemental insurance on hospital admission is also not significant in previous years (Panel B of table 7). These findings are in line with other studies which find no statistical significant incentive effect on the probability of any hospital admission (Deb and Trivedi [15], Manning *et al.* [35]).

A glance at the tables reveals also that: (i) previous hospital staying has a positive and statistical significant residual effect on the probability of having any hospital admission (which may be related to the existence of preconditions in individual health); (ii) past insurance status has a positive and statistically significant effect on insurance choice; (iii) the utilization of any previous hospital admission is negligible and statistically not significant (see Panel A and B of table 7). The strong state dependence in insurance choice (β^{S_2} and β^{S_3} are generally greater than 3) may occur because of the existence of a learning effect which may be linked to individuals cognitive abilities which have been found an important factor of selection in Medigap insurance (Fang et al. [20]).

Results under different specifications do not differ substantially also for the estimated effect of other controls on insurance choice and hospital utilization at year 2006. Table 8 reports the δ^{S_3} and δ^{M_3} coefficients. Individuals in the top quartile of the wealth distribution are more likely to subscribe additional coverage but less to have hospital inpatient staying. People covered by other public insurance program - such as Medicaid or Tricare - purchase less additional health insurance coverage but use more resources. On the contrary individual covered by additional coverage purchased by a spouse or a former employer in the previous year are more likely to subscribe supplemental insurance although the effect on utilization is not statistically significant.

Let us consider now the random intercepts α which describe the effect of U on the responses. Results of equation (12) are organized as follows. Table 4 reports the class membership probabilities (α^U) and the estimated intercepts for each class of the equations describing the health status (α^{H_1}), the investment decision on risk reducing behavior: smoking (α^{H_2}) and preventive care (α^{H_3}); table 5, 6, 7 and 8 show the estimated parameters of the equations describing respectively the insurance choice and the hospital admission over the three waves.

In the sequel, for the sake of brevity we will comment on these estimated coefficients for the case of four LC ($m = 4$).

Table 4 shows the existence of two groups related to health status: individuals in class three ($U = 3$) and four ($U = 4$) have on average 60% more probability to have a good health status than those in class one ($U = 1$) and two ($U = 2$). Risk preferences indicators show a much more complex picture. Individuals' types two and four tend to be much more cautious than people of types one or three. In fact they tend to be non-smoker and undertake more than the 50% of gender-appropriate preventive health care measures in the previous two years. From table 5 types two and four are more likely to take out supplemental insurance and show an higher propensity to be continuously covered by additional plan over the time. In particular those individual are 20% more likely to purchase supplemental insurance than all the others. Finally looking at table 6 people in class one and two are more likely to have any hospital admission over the time than individuals in class three or four.

The above results suggest that the four "types" have the following characteristics: type three individuals are unfrequent users with a good health status as type four, but differently from the latter they have low preference for cautiousness. Those of type two have a worse health status (high actual risk) and show an higher propensity to invest in insurance coverage and risk reducing activities than those in class one. Therefore different types purchase insurance motivated by different reasons: high actual risk (types one and two) and low risk tolerance (types three and four). Ex post, people in the first two groups are higher risk than "type" four individuals which use less hospital resources than all the others.

These findings confirm the multidimensional nature of unobserved heterogeneity so forcibly stressed by Finkelstein and McGarry [22]. Recall that if there was a unique variable which underlies unobserved heterogeneity (e.g. actual riskness), and this variable was monotonically related to all the observed responses, then the random effects coefficients would have the same order across classes in each of the observed responses. As seen before, this is not true in our sample, since there are many instances where the probability of success is higher for a certain type than another for a given observed response but the opposite holds for another one (just as an example, for $m = 4$ we have $\alpha_1^{S_3} < \alpha_4^{S_3} < \alpha_3^{S_3} < \alpha_2^{S_3}$ but $\alpha_3^{M_3} < \alpha_4^{M_3} < \alpha_2^{M_3} < \alpha_1^{M_3}$). The question then naturally arises is whether the observed multidimensional pattern of U is simply due to sampling variation, or rather to the presence of more than one underlying unobservable variable which pull U in different directions.

As discussed above in section 4, techniques of order restricted inference can be employed to test the monotonicity assumption for the standard textbook insurance model; since U has four levels, and since the four random effects are ordered such that health status is increasing in U, the assumption that the conditional probability of insurance and hospital admission responses is monotonically decreasing in U involves imposing 9 inequality constraints. The LR test statistic of this hypothesis is equal to 72.5. The conservative 5% critical value (Kodde and Palm ([32], page 1246) is equal to 16.274; thus the non monotonic pattern of insurance purchase and hospital suggests the presence of more than one underlying unobservable variables which have contrasting effects on the decision to purchase insurance and use health care.[8]

7. Final remarks

In the health insurance market consumers have private information about their health status (actual risk) and preferences. To deal with this source of heterogeneity the traditional approach relies on bivariate probit models with observable proxies. A common finding in empirical studies based on the bivariate probit is i) a significant effect of supplemental insurance on health care measure and ii) a negative correlation between risk occurrence measure and insurance coverage. In this paper, we exploit some recent developments in the latent class modeling to explore the extent to which supplemental health insurance (Medigap) affects hospital utilization after controlling for unobservable characteristics which drive the decision to purchase health insurance as well as the probability to risk occurrence. Our model is motivated by the existence of selection in the Medigap market and multiple dimension of private information ((Finkelstein [21]) , Finkelstein and McGarry [22], Cutler *et al.* [11] and Fang *et al.* [20]). To this purpose we identify unobserved individual "types" representing a mix between health status (actual risk), cautiousness, and individual preferences for insurance purchase and health care utilization. Our main findings in the HRS sample are: i) there seem to be no evidence of incentive effect of supplemental Medigap insurance on hospital admission; ii) there is a substantial state dependence in the supplementary health insurance decisions; iii) there is a substantial unobserved heterogeneity which drives health insurance purchase and hospital use; iv) the unobserved heterogeneity seem to be generated by multiple sources.

[8]The LR test statistic for 3 and 5 LC is respectively given to 48.6 and 78.4, with 5% critical values respectively equal to 9.99 and 17.94, thus the standard unidimensional insurance model is overwhelmingly rejected in all specification.

Appendix A. Estimation and Identification of model (12)

A.1. Likelihood.

A.1.1. *Incomplete data (observable) likelihood.* Let $\boldsymbol{Y}$ denote the 9-dimensional vector of observable binary response variables employed in model (12), that is

$$\boldsymbol{Y} = [M_t, S_t, M_{t-1}, S_{t-1}, M_{t-2}, S_{t-2}, H_1, H_2, H_3]$$

and let $\boldsymbol{y}$ be any of the $t = 2^9$ possible response configuration. Let then $q_{\boldsymbol{y}} = Pr(Y_1 = y_1, \ldots Y_K = y_K)$, denote the corresponding cell probabilities, which may be arranged into the t-sized vector $\boldsymbol{q}$ lexicographically by letting the elements of $\boldsymbol{y}$ with a larger index run faster.

Assume we have observations on $\boldsymbol{Y}$ for n individuals. For subjects with covariates $\boldsymbol{x}_i$, let

- $\boldsymbol{y}(i)$ be the response configuration,
- $\boldsymbol{n}(i)$ be the t-sized vector containing the frequency table of $\boldsymbol{Y}$ in lexicographic order; if there is a single subject with covariates $\boldsymbol{x}_i$ (for example when $\boldsymbol{x}$ contains continuous covariates), then $\boldsymbol{n}(i)$ is a vector of zeros except for a one in the cell corresponding to the response pattern $\boldsymbol{y}(i)$;
- $\boldsymbol{q}(i)$ be the t-sized vector denoting the probability distribution.

The (kernel of the) log-likelihood may be written as

$$L = \sum L_i = \sum \boldsymbol{n}(i)' \ln[\boldsymbol{q}(i)].$$

A.1.2. *Complete data (unobservable) likelihood.* Consider now the 10-sized unobservable vector $[U, \boldsymbol{Y}]$, $u, \boldsymbol{y}$ be any of the possible response configuration. Let then $p_{u,\boldsymbol{y}} = Pr(U = u, Y_1 = y_1, \ldots Y_K = y_K)$, denote the corresponding cell probabilities, which may be arranged into the $m \times t$-sized vector $\boldsymbol{p}$ lexicographically. by letting the elements of $\boldsymbol{y}$ with a larger index run faster. Let also $\boldsymbol{L} = (\boldsymbol{1}_m{}' \otimes \boldsymbol{I}_{2^9})$ denote the matrix which marginalizes with respect to the latent variable U,

For subjects with covariates configuration $\boldsymbol{x}_i$ let

- $\boldsymbol{p}(i)$ be the $m \times t$-sized vector containing the joint probability distribution of $(U, \boldsymbol{Y})$;
- $\boldsymbol{m}(i)$ the $m \times t$-sized vector containing the unobservable frequency table of $(U, \boldsymbol{Y})$,

so that $\boldsymbol{n}(i) = \boldsymbol{L}\boldsymbol{m}(i)$ and $\boldsymbol{q}(i) = \boldsymbol{L}\boldsymbol{p}(i)$. If the latent class U could be observed, the corresponding (kernel of the) log-likelihood would have the form

$$\Lambda = \sum \Lambda_i = \sum \boldsymbol{m}(i)' \ln[\boldsymbol{p}(i)].$$

A.2. EM algorithm.

Maximizing the log-likelihood is as a problem of incomplete data which may be tackled by the EM algorithm (Dempster *et al.* [17]).

A.2.1. *The E step.* Since the multinomial is a member of the exponential family, the conditional expectation involved in the E step is equivalent to computing the posterior probability of latent class U given the observed configuration $\boldsymbol{y}$, so that $m_{u,\boldsymbol{y}}(i) = n_{\boldsymbol{y}}(i)Pr(U = u \mid \boldsymbol{y}_i, \boldsymbol{x}_i)$ follows from a simple expectation of a multinomial distribution for U.

A.2.2. *The M step.* Implementation of the method of scoring for the maximization of Λ with respect to the model parameters $\boldsymbol{\psi}$ requires computation of the score vector (first derivative with respect to $\boldsymbol{\psi}$) and of the expected information matrix (minus the expected value of the second derivative). Since Λ is a multinomial log-likelihood, exponential family results can be exploited to make such calculations straightforward. In practice, after rewriting Λ in terms of the canonical parameters of the multinomial distribution, say $\boldsymbol{\theta}(i)$, there are invertible and differentiable mappings from $\boldsymbol{\theta}(i)$ to the vector of probabilities $\boldsymbol{p}(i)$ and from $\boldsymbol{p}(i)$ to $\boldsymbol{\lambda}(i)$ (the latter mapping is described in Lang [34] and Bartolucci *et al.* [4]), while $\boldsymbol{\lambda}(i)$ is linked to $\boldsymbol{\psi}$ by the linear regression model (12). The interested reader may see Dardanoni and Forcina [13] for details.

A.3. **Estimation of the variance matrix.** While the EM algorithm is a very robust method of estimation of the model parameters in the presence of unobservables, it does not provide a consistent estimate of the variance matrix of the model parameters, since the expected information matrix of the complete data likelihood is based on the assumption that $\boldsymbol{m}$ is known. Thus, using its inverse as an estimate of the variance matrix implies that standard errors will generally be underestimated.

The correct information matrix may be computed from the complete data log-likelihood as follows. Write $L_i = \boldsymbol{n}(i)'\tilde{\boldsymbol{G}}\boldsymbol{\gamma}_i - n_i \ln[\boldsymbol{1}' \exp(\tilde{\boldsymbol{G}}\boldsymbol{\gamma}_i)]$ where $\boldsymbol{\gamma}_i$, the canonical parameter of the observed multinomial, may be written as $\tilde{\boldsymbol{H}} \ln[\boldsymbol{L} \exp(\boldsymbol{G\theta}_i)/\boldsymbol{1}' \exp(\boldsymbol{G\theta}_i)]$, where $\tilde{\boldsymbol{H}}$ is a $t \times (t-1)$ contrast matrix used to define the canonical parameters and $\tilde{\boldsymbol{G}}$ is its right inverse while $\boldsymbol{G}$ is the design matrix which defines the canonical parameters $\boldsymbol{\theta}$ for the latent distribution $\boldsymbol{p}(i)$ which has v columns of full rank. By differentiating L_i by the chain rule with respect to $\boldsymbol{\psi}$ one may write

$$\sum \frac{\partial L_i}{\partial \boldsymbol{\psi}} = \boldsymbol{B}_i{}' \boldsymbol{R}_i{}' \boldsymbol{G}' \boldsymbol{\Omega}_i \boldsymbol{L}' diag(\boldsymbol{q}_i)^{-1} \tilde{\boldsymbol{H}}' \tilde{\boldsymbol{G}}' (\boldsymbol{n}(i) - n_i \boldsymbol{q}_i)$$

where $n_i = \boldsymbol{1}'\boldsymbol{n}(i)$, $\boldsymbol{\Omega}_i = diag[\boldsymbol{p}(i) - \boldsymbol{p}(i)\boldsymbol{p}(i)']$ and $\boldsymbol{R}_i$ is the derivative of the canonical parameter $\boldsymbol{\theta}_i$ with respect to $\boldsymbol{\lambda}_i{}'$. Because $E(\boldsymbol{n}(i) - n_i\boldsymbol{q}_i) = \boldsymbol{0}$, minus the expected value of the second derivative may be written as

$$\boldsymbol{F}_i = \boldsymbol{B}_i{}' \boldsymbol{R}_i{}' \boldsymbol{G}' \boldsymbol{\Omega}_i \boldsymbol{L}' diag(\boldsymbol{q}_i)^{-1} \tilde{\boldsymbol{H}}' \tilde{\boldsymbol{G}}' \boldsymbol{L} \boldsymbol{\Omega}_i \boldsymbol{G} \boldsymbol{R}_i \boldsymbol{B}_i$$

(where $\tilde{\boldsymbol{H}}'\tilde{\boldsymbol{G}}'$ is simply equal to $\boldsymbol{I}_t - \boldsymbol{1}_t\boldsymbol{1}_t{}'/t$), so the information matrix is simply $\sum_i \boldsymbol{F}_i$. See also Forcina [23].

A.4. **Model identifiability.** Formally identifiability concerns the mapping from the manifest probability distribution $\boldsymbol{q}$ and the model parameter $\boldsymbol{\psi}$ (Rothemberg, [40]). Recall that a model is locally identified when at any $\boldsymbol{\psi}_0$ the set of points such that $\|\boldsymbol{q}(\boldsymbol{\psi}) - \boldsymbol{q}(\boldsymbol{\psi}_0)\| = 0$ satisfy $\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\| > \delta > 0$. This implies that there exist no parameter value with a neighborhood where the likelihood is constant and thus the information matrix must be positive definite everywhere.

To analyse local identifiability of our model, following Forcina [23], let $\boldsymbol{\gamma}$ denote the vector obtained by stacking the vectors $\boldsymbol{\gamma}_i$ (the vectors of canonical parameters of the saturated log-linear model for each subject in the manifest distribution) and consider the different parametric transformations involved:

(1) from $\boldsymbol{\gamma}$ to $\boldsymbol{\theta}$, the vector obtained by stacking the vectors of canonical parameters of the latent class model for each subject,
(2) from $\boldsymbol{\theta}$ to $\boldsymbol{\lambda}$, the vector obtained by stacking the vectors of marginal parameters for each subject,
(3) the regression model $\boldsymbol{\lambda} = \boldsymbol{B}\boldsymbol{\psi}$.

Identifiability of the regression model is easily established by checking that $\boldsymbol{B}$ is of full column rank. Results from Bartolucci *et al.* [4] (Theorem 1 p.703) ensure that the transformation to the marginal parameters is invertible and differentiable. So, the crucial transformation is the first one. Though no analytic result is available to check the full rank of the jacobian of the mapping from $\boldsymbol{q}$ to $\boldsymbol{\psi}$, this may be tested numerically in a fast and efficient way as described by Forcina [23] for a wide range of values sampled at random. Since in our case in 10000 runs no instance was detected where the rank of the jacobian was any close to being deficient, we may conclude that our model is identifiable for a wide range of parameter values.

Appendix B. Parameters' estimates

Table 1. Sample Characteristics and Variable Definition

Variable	Definition of Binary Variables	Full Sample	No Insurance	With Insurance
Insurance Status				
spins06	1 = enrolled in Medigap at 2008.	0.33	-	-
spins04	1 = enrolled in Medigap at 2004.	0.35	0.15	0.75
spins02	1 = enrolled in Medigap at 2002.	0.33	0.16	0.67
medid06	1 = covered by Medicaid.	0.13	0.18	0.03
tri06	1 = covered by Tri-care.	0.09	0.11	0.04
iemp04	1 = additional coverage from former employer at 2004.	0.08	0.09	0.08
iemps04	1 = additional coverage from spouse at 2004.	0.04	0.04	0.04
Health Care				
h06	1 = entered a hospital in the preceding two years from 2006.	0.36	0.36	0.38
h04	1 = entered a hospital in the preceding two years from 2004.	0.33	0.33	0.31
h02	1 = entered a hospital in the preceding two years from 2002.	0.30	0.29	0.31
Demographics				
age 70-80	1 = aged between 70 and 80 years.	0.55	0.56	0.53
age 80-90	1 = aged between 80 and 90 years.	0.24	0.23	0.26
age > 90	1 = older than 90 years.	0.04	0.04	0.05
fem	1 = female.	0.48	0.48	0.49
hedu	1 = highest educational attainment college or above.	0.33	0.30	0.39
ass06q4	1 = if individual is in the top wealth quartile.	0.22	0.18	0.32
ass06q3	1 = if individual is in the third wealth quartile.	0.22	0.21	0.25
ass06q2	1 = if individual is in the second wealth quartile.	0.26	0.26	0.26
Health Status				
he06	1 = at least two good health measures among ADL, IADL, disease and mobility limitation.	0.39	0.37	0.42
Risk Behavior				
smoken06	1 = current smoker.	0.15	0.16	0.11
prev	1 = more than 50% of sex adjusted preventive care.	0.79	0.77	0.83

Table 2. Probit for Hospital Admission and Insurance Choice at 2006

	Probit Model		Bivariate Probit Model	
Independent Variables	Hospital 2006	Insurance 2006	Hospital 2006	Insurance 2006
spins06	0.18**	.	0.94**	.
	(0.07)	.	(0.32)	.
spins04	-0.07	1.38**	-0.41**	1.38**
	(0.08)	(0.07)	(0.16)	(0.07)
spins02	0.01	0.71**	-0.16	0.71**
	(0.07)	(0.07)	(0.10)	(0.07)
age 70-80	0.09	-0.08	0.10	-0.06
	(0.07)	(0.08)	(0.07)	(0.08)
age 80-90	0.20*	0.03	0.19*	0.07
	(0.08)	(0.09)	(0.08)	(0.10)
age $>$ 90	0.15	0.12	0.12	0.14
	(0.14)	(0.17)	(0.14)	(0.17)
fem	-0.11*	0.06	-0.11*	0.07
	(0.05)	(0.06)	(0.05)	(0.06)
hedu	0.09	0.04	0.08	0.04
	(0.06)	(0.07)	(0.06)	(0.06)
ass06q4	-0.22**	0.30**	-0.27**	0.31**
	(0.08)	(0.10)	(0.08)	(0.10)
ass06q3	-0.1	0.18*	-0.13	0.19*
	(0.08)	(0.09)	(0.08)	(0.09)
ass06q2	-0.15*	0.13	-0.16*	0.13
	(0.07)	(0.08)	(0.07)	(0.08)
he06	-0.51**	0.02	-0.5**	0.02
	(0.06)	(0.06)	(0.06)	(0.06)
h04	0.59**	-0.07	0.58**	-0.07
	(0.05)	(0.07)	(0.06)	(0.07)
h02	0.19**	0.04	0.18**	0.04
	(0.06)	(0.07)	(0.06)	(0.06)
medid06	0.10	-0.35**	0.13	-0.34**
	(0.08)	(0.12)	(0.08)	(0.12)
tri06	0.05	-0.59**	0.13	-0.59**
	(0.09)	(0.12)	(0.09)	(0.12)
iemp04	0.01	0.61**	-0.11	0.63**
	(0.09)	(0.10)	(0.10)	(0.10)
iemps04	0.07	0.65**	-0.05	0.65**
	(0.13)	(0.14)	(0.14)	(0.14)
smoken06	-0.25*	-0.03	-0.23**	-0.02
	(0.08)	(0.09)	(0.07)	(0.09)
prev	0.05	0.02	0.05	0.03
	(0.06)	(0.07)	(0.06)	(0.07)
constant	-0.47**	-1.48**	-0.50**	-1.52**
	(0.11)	(0.13)	(0.11)	(0.13)
ρ	.-	.	-0.43	.
	.	.	(0.18)	.
# of Obs.	2931	2934	2931	.
Log-likelihood	-1730.25	-1209.39	-2937.63	.
Pseudo R^2	0.10	0.35	.	.

Note: Standard errors are reported in brackets
** Significant at the 1% level;
* Significant at the 5% level.

Table 3. Model Selection Criteria for system (12)

	Number of Latent Classes			
	2LC	3LC	4LC	5LC
$L(\psi)$	-14326	-14268	-14229	-14221
$BIC(\psi)$	29069	29031	29033	29097
$AIC(\psi)$	28757	28660	28603	28606

Table 4. Estimated intercepts α for system (12): risk reducing behavior and health

	2LC			3LC			4LC			5LC		
	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.
Class Memb. Pr.												
α_2^U	0.05	0.18	0.51	-0.31	0.24	0.27	-0.37	0.20	0.22	-0.54	0.23	0.18
α_3^U	.	.	.	0.29	0.22	0.36	-0.15	0.29	0.19	-0.66	0.58	0.10
α_4^U	.	.	.	.	.	.	0.23	0.42	0.24	0.62	0.68	0.17
α_5^U	.	.	.	.	.	.	.	.	.	0.32	0.47	0.24
Health Status												
$\alpha_1^{H_1}$	-1.48	0.14	0.18	-1.82	0.32	0.13	-2.07	0.39	0.11	-1.86	0.31	0.13
$\alpha_2^{H_1}$	0.32	0.13	0.57	-1.09	0.15	0.25	-1.39	0.21	0.19	-1.73	0.34	0.15
$\alpha_3^{H_1}$	.	.	.	1.05	0.36	0.74	0.33	0.20	0.58	-0.42	0.34	0.39
$\alpha_4^{H_1}$	.	.	.	.	.	.	1.24	0.47	0.77	0.74	0.32	0.67
$\alpha_5^{H_1}$	.	.	.	.	.	.	.	.	.	0.77	0.33	0.68
Cur. Smoker												
$\alpha_1^{H_2}$	-2.07	0.11	0.11	-1.60	0.11	0.16	-1.77	0.14	0.14	-1.84	0.16	0.13
$\alpha_2^{H_2}$	-1.47	0.08	0.18	-2.40	0.21	0.08	-2.25	0.20	0.09	-2.66	0.38	0.06
$\alpha_3^{H_2}$	.	.	.	-1.80	0.10	0.14	-0.69	0.21	0.33	-0.11	0.50	0.47
$\alpha_4^{H_2}$	.	.	.	.	.	.	-2.74	0.51	0.06	-1.26	0.23	0.21
$\alpha_5^{H_2}$	.	.	.	.	.	.	.	.	.	-2.88	0.59	0.05
Prev. Care												
$\alpha_1^{H_3}$	1.40	0.08	0.80	0.98	0.09	0.72	1.13	0.12	0.75	1.13	0.12	0.75
$\alpha_2^{H_3}$	1.17	0.07	0.76	1.75	0.15	0.85	1.57	0.16	0.82	1.61	0.21	0.83
$\alpha_3^{H_3}$	.	.	.	1.30	0.10	0.78	0.21	0.25	0.55	-0.07	0.43	0.48
$\alpha_4^{H_3}$	.	.	.	.	.	.	3.15	1.10	0.95	0.78	0.23	0.68
$\alpha_5^{H_3}$	.	.	.	.	.	.	.	.	.	3.35	1.26	0.96

Table 5. Estimated intercepts α of system (12): insurance choice

	2LC			3LC			4LC			5LC		
	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.
Sup. Ins. 02												
$\alpha_1^{S_1}$	-0.58	0.07	0.35	-1.93	0.28	0.12	-1.81	0.24	0.14	-1.93	0.27	0.12
$\alpha_2^{S_1}$	-0.83	0.07	0.30	0.67	0.25	0.66	0.74	0.24	0.67	0.71	0.26	0.67
$\alpha_3^{S_1}$	.	.	.	-0.92	0.11	0.28	-1.48	0.25	0.18	-0.60	0.35	0.35
$\alpha_4^{S_1}$	.	.	.	.	.	.	-0.53	0.15	0.37	-1.96	0.53	0.12
$\alpha_5^{S_1}$	.	.	.	.	.	.	.	.	.	-0.12	0.21	0.46
Sup. Ins. 04												
$\alpha_1^{S_2}$	-2.00	0.14	0.11	-1.82	0.15	0.13	-1.65	0.16	0.16	-1.70	0.16	0.15
$\alpha_2^{S_2}$	-1.62	0.09	0.16	-2.06	0.34	0.11	-2.30	0.37	0.09	-2.45	0.42	0.07
$\alpha_3^{S_2}$	.	.	.	-1.64	0.12	0.16	-2.43	0.32	0.08	-2.81	0.53	0.05
$\alpha_4^{S_2}$	.	.	.	.	.	.	-1.25	0.18	0.22	-1.83	0.27	0.13
$\alpha_5^{S_2}$	.	.	.	.	.	.	.	.	.	-1.33	0.25	0.20
Sup. Ins. 06												
$\alpha_1^{S_3}$	-2.13	0.23	0.10	-3.76	0.49	0.02	-4.06	0.52	0.01	-4.06	0.52	0.01
$\alpha_2^{S_3}$	-2.61	0.21	0.06	-0.64	0.37	0.34	-0.50	0.37	0.37	-0.46	0.40	0.38
$\alpha_3^{S_3}$	.	.	.	-2.58	0.28	0.07	-2.54	0.31	0.07	-1.61	0.46	0.16
$\alpha_4^{S_3}$	.	.	.	.	.	.	-2.84	0.38	0.05	-3.20	0.46	0.03
$\alpha_5^{S_3}$	.	.	.	.	.	.	.	.	.	-2.47	0.37	0.07

Table 6. Estimated intercepts α of system (12): hospital admission

	2LC			3LC			4LC			5LC		
	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.	Coef.	St.Er.	Prob.
Hosp. Adm. 02												
$\alpha_1^{M_1}$	-0.31	0.09	0.42	-0.53	0.11	0.37	-0.41	0.11	0.39	-0.43	0.12	0.39
$\alpha_2^{M_1}$	-1.55	0.12	0.17	-0.39	0.19	0.40	-0.36	0.19	0.40	-0.18	0.21	0.45
$\alpha_3^{M_1}$	.	.	.	-1.73	0.16	0.14	-2.45	0.45	0.07	-0.90	0.38	0.28
$\alpha_4^{M_1}$	.	.	.	.	.	.	-1.11	0.17	0.24	-7.48	48.27	0.00
$\alpha_5^{M_1}$	.	.	.	.	.	.	.	.	.	-0.73	0.24	0.32
Hosp. Adm. 04												
$\alpha_1^{M_2}$	-0.10	0.18	0.47	-0.61	0.14	0.34	-0.44	0.15	0.39	-0.37	0.16	0.40
$\alpha_2^{M_2}$	-2.04	0.19	0.11	-0.32	0.18	0.42	-0.25	0.18	0.43	-0.03	0.21	0.49
$\alpha_3^{M_2}$	.	.	.	-2.00	0.20	0.11	-2.25	0.34	0.09	-1.30	0.39	0.21
$\alpha_4^{M_2}$	.	.	.	.	.	.	-1.57	0.20	0.17	-3.27	1.19	0.03
$\alpha_5^{M_2}$	.	.	.	.	.	.	.	.	.	-1.31	0.24	0.21
Hosp. Adm. 06												
$\alpha_1^{M_3}$	0.25	0.33	0.56	-0.39	0.22	0.40	-0.25	0.22	0.44	-0.12	0.24	0.46
$\alpha_2^{M_3}$	-2.02	0.26	0.11	-0.40	0.27	0.39	-0.26	0.27	0.43	0.02	0.34	0.50
$\alpha_3^{M_3}$	.	.	.	-2.14	0.28	0.10	-2.18	0.35	0.10	-2.16	0.69	0.10
$\alpha_4^{M_3}$	.	.	.	.	.	.	-1.67	0.25	0.15	-2.04	0.35	0.11
$\alpha_5^{M_3}$	.	.	.	.	.	.	.	.	.	-1.64	0.30	0.16

Table 7. Estimated β and γ parameters of system (12)

	2LC		3LC		4LC		5LC	
	Coef.	St.Er.	Coef.	St.Er.	Coef.	St.Er.	Coef.	St.Er.
Panel A: eq. (9)								
Sup. Ins. 06								
β^{S_3}	2.91**	0.10	3.07**	0.22	3.32**	0.28	3.29**	0.27
γ^{S_3}	-0.25	0.15	-0.42	0.23	-0.33	0.21	-0.40	0.21
Hosp. Adm. 6								
β^{M_3}	0.12	0.26	0.16	0.17	0.14	0.16	0.22	0.18
γ^{M_3}	0.24	0.13	0.72**	0.13	0.66**	0.13	0.60**	0.15
Panel B: eq. (11)								
Sup. Ins. 04								
β^{S_2}	3.01**	0.10	3.09**	0.21	3.22**	0.13	3.26**	0.26
γ^{S_2}	0.09	0.12	0.06	0.12	0.05	0.11	0.02	0.13
Hosp. Adm. 02								
β^{M_1}	0.06	0.09	-0.01	0.14	-0.03	0.13	-0.18	0.16
Hosp. Adm. 04								
β^{M_2}	-0.11	0.11	-0.22*	0.11	-0.25*	0.12	-0.23	0.12
γ^{M_2}	0.71**	0.12	0.84**	0.10	0.81**	0.11	0.71**	0.12

** Significant at the 1% level;
* Significant at the 5% level.

Table 8. Estimated δ parameters for system (12)

	2LC		3LC		4LC		5LC	
	Coef.	St. Er.	Coef.	St. Er.	Coef.	St. Er.	Coef.	St. Er.
Sup. Ins. 2006								
age 70-80	-0.03	0.13	-0.12	0.16	-0.13	0.17	-0.13	0.16
age 80-90	0.11	0.15	0.07	0.19	0.07	0.19	0.07	0.19
age > 90	0.24	0.27	0.27	0.33	0.27	0.34	0.24	0.34
fem	0.12	0.10	0.15	0.12	0.11	0.12	0.11	0.12
hedu	0.06	0.11	0.10	0.14	0.15	0.14	0.15	0.14
ass06q2	0.33*	0.14	0.22	0.17	0.22	0.17	0.23	0.17
ass06q3	0.41**	0.15	0.31	0.18	0.34	0.19	0.36*	0.18
ass06q4	0.71**	0.15	0.52*	0.19	0.56**	0.19	0.59*	0.19
medid06	-0.75**	0.17	-0.77**	0.21	-0.79*	0.22	-0.80**	0.21
tri06	-1.02**	0.19	-1.28**	0.25	-1.27**	0.25	-1.23**	0.24
iemp04	1.09**	0.17	1.30**	0.22	1.37**	0.23	1.36**	0.23
iemps04	1.08	0.23	1.44**	0.30	1.48**	0.31	1.48**	0.31
Hosp. Adm. 2006								
age 70-80	0.19	0.13	0.17	0.12	0.16	0.12	0.16	0.12
age 80-90	0.43*	0.15	0.38*	0.13	0.37*	0.14	0.37*	0.14
age > 90	0.35	0.27	0.29	0.24	0.27	0.24	0.25	0.25
fem	-0.19	0.10	-0.20*	0.09	-0.19*	0.09	-0.20*	0.09
hedu	0.16	0.11	0.15	0.10	0.14	0.10	0.16	0.10
ass06q2	-0.26	0.14	-0.23	0.12	-0.23	0.12	-0.26*	0.13
ass06q3	-0.17	0.14	-0.15	0.13	-0.17	0.13	-0.18	0.13
ass06q4	-0.39*	0.15	-0.34*	0.13	-0.35*	0.14	-0.38*	0.14
medid06	0.19	0.16	0.17	0.14	0.21	0.14	0.21	0.14
tri06	0.09	0.17	0.09	0.15	0.10	0.15	0.10	0.16
iemp04	0.03	0.18	0.07	0.16	0.07	0.16	0.04	0.16
iemps04	0.13	0.25	0.15	0.22	0.16	0.22	0.13	0.23

** Significant at the 1% level;
* Significant at the 5% level.

References

[1] Arrow, K. J. (1963): "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53(5), pp. 485-973.
[2] Bago d'Uva, T. and Santos Silva J. M. C. (2002): "Asymmetric Information in the Portuguese Health Insurance Market", ISEG, Universidade Técnica de Lisboa, *mimeo.*
[3] Bartolucci, F. and Forcina, A. (2006): "A class of latent marginal models for capture-recapture data with continuous covariates," *Journal of the American Statistical Association*, 101, pp. 786-794.
[4] Bartolucci, F., Colombi, R. and Forcina, A. (2007): "An extended class of marginal link functions for modelling contingency tables by equality and inequality constraints," *Statistica Sinica*, 17, pp. 691-711.
[5] Buchmueller, T.C., Couffinhal, A., Grignon, M. and Perronnin, M. (2004): "Access to physician services: does supplemental insurance matter? Evidence for France," *Health Economics*, 13, pp. 669-687.
[6] Cameron, A.C. and P.K. Trivedi (1991): "The role of income and health risk in the choice of health insurance," *Journal of Public Economics*, 45, pp. 1-28.
[7] Cameron, A.C., Trivedi, P.K., Milne F., and Piggott J. (1988): "A microeconometric model of demand for health care and health insurance in Australia," *Review of Economic Studies*, 55, pp. 85-106.
[8] Chiappori P. A., Durand F. and Geoffard, P. Y. (1998): "Moral hazard and the demand for physician services: first lessons from a French natural experiment," *European Economic Review*, 42, pp. 499-511.
[9] Chiappori, P. A. (2000): "Econometric Models of Insurance under Asymmetric Information," In Georges Dionne, ed. *Handbook of Insurance Economics*. London: Kluwer, pp. 365-394.
[10] Chiappori, P. A. and Salaniè, B. (1997): "Empirical contract theory: The case of insurance data," *JEuropean Economic Review*, 41(3-5), pp. 943-950.
[11] Cutler, D. M., Finkelstein, A. and McGarry, K. (2008): "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," *NBER Working Paper*, 13746.
[12] Dardanoni, V., Forcina, A., (1998): "A Unified Approach to Likelihood Inference on Stochastic Orderings in a Nonparametric Context." *Journal of the American Statistical Association*, 93, pp. 1112-1123.
[13] Dardanoni, V. and A. Forcina (2008): "Multivariate ordered regression", *mimeo*
[14] Dardanoni, V., Forcina, A., and Modica, S. (2008): "Direct causal effects in education transmission," *submitted.*
[15] Deb, P. and Trivedi, P. K. (1997):"Demand for medical care by the elderly: a finite mixture approach," *Journal of Applied Econometrics*, 12, pp. 313-336.
[16] Deb, P. and Trivedi, P. K. (2002): "The structure of demand for health care: latent class versus two-part models," *Journal of Health Economics*, 21(4), pp. 601-625.
[17] Dempster, A. P., N. M. Laird and D. B. Rubin (1977): "Maximum Likelihood for Incomplete Data Via the EM Algorithm", *Journal of the Royal Statistical Society* Series B **39**, pp. 1–22
[18] Eichner, M. J. (1998): "The demand for medical care: What people pay does matter," *American Economic Review*, 88(2), pp. 117-121.
[19] Ettner, S, (1997): "Adverse Selection and the Purchase of Medigap Insurance by the Elderly," *Journal of Health Economics*, 16(5), pp. 499-624.
[20] Fang, H., Keane, M.P. and Silverman, D., (2008) "Sources of Advantageous Se- lection: Evidence from the Medigap Insurance Marke," Journal of Political Economy 116, p.303-350.
[21] Finkelstein, A. (2004): "Minimum standards, insurance regulation and adverse selection: evidence from the Medigap market," *Journal of Public Economics*, 88(12), pp. 2515-2547.
[22] Finkelstein, A. and McGarry, K. (2006): "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *The American Economic Review*, 4, pp. 938-958.
[23] Forcina, A. (2008): "Identifiability of Extended Latent Class Models with Individual Covariates," *Computational Statistics and Data Analysis*, 52, pp. 5263-5268
[24] Gourieroux, C. and Monfort, A. (1995): "Statistics and Econometrics Models", Cambridge: *Cambridge University Press.*
[25] Heckman, J. J. and Singer, B. (1984): "Econometric duration analysis," *Journal of Econometrics*, 24(1-2), pp. 63-132.
[26] Holly, A., Gardiol L., Domenighetti, G. and Bisig, B. (1998): "An econometric model of health care utilization and health insurance in Switzerland," *European Economic Review*, 42, pp. 513-522.
[27] Huang G., Bandeen-Roche K. (2004): "Building an identifiable latent variable model with covariate effects on underlying and measured variables," *Psychometrika* 69(1), pp. 5-32.

[28] Hurd, M. and McGarry, K. (1997): "Medical Insurance and the Use of Health Care Services by the Elderly," *Journal of Health Economics*, 16(2), pp. 129-154.
[29] Jones, A.M., Koolman, X. and Van Doorslaer, E. (2006), "The impact of supplementary private health insurance on the use of specialists in selected European countries," *Annales d'Economie et de Statistique*, 83-84, pp. 251-275.
[30] Khandker, R. K. and McCormack,L. A. (1996): "Enrollment and Utilization across Medicare Supplemental Plans." Report Submitted to the Health Care Financing Administration by the Center for Health Economics Research.
[31] Khandker, R. K. and McCormack,L. A. (1999): "Medical Spending by Beneficiaries with Various Types of Supplemental Insurance," *Medical Care Research and Review*, 56, pp. 137-155.
[32] Kodde, D. A. and Palm, F. C. (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions", *Econometrica*, 54(5), pp. 1243-1248.
[33] Landerman, L., Fillerbaum, G., Pieper, C., Maddox, G., Gold D. and Guralnik, J. (1998): "Private Health Insurance Coverage and Disability Among Older Americans," *Journal of Gerontology*, 53, pp. S258-S266.
[34] Lang, J. B. (1996): "Maximum likelihood methods for a generalized class of log-linear models," *Annals of statistics*, 24, pp. 726-752.
[35] Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B. and Leibowitz, A. (1987): "Health Insurance and the Demand for Medical Care: Evidence from a randomized experiment," *American Economic Review*, 77, pp. 251-277.
[36] Pauly, M.V. (1968): "The economics of moral hazard", *American Economic Review*, 58, pp. 533-539.
[37] Propper, C. (1989): "An econometric analysis of the demand for private insurance in England and Wales," *Applied Economics*, 21, pp. 777-792.
[38] Propper, C. (1993): "Constrained Choice Sets in the UK Demand for Private Medical Insurance," *Journal of Public Economics*, 51, pp.287-307.
[39] Rand HRS Data (Version H). Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (2008).
[40] Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, **39**, pp. 577-591.
[41] Vera-Hernàndez, A.M. (1999): "Duplicate coverage and the demand for health care: the case of Catalonia," *Health Economics*, 8, pp. 579-598.
[42] Vistnes, J. and Banthin, J. (1998): "The Demand for Medicare Supplemental Insurance Benefits: The Role of Attitudes toward Medical Care and Risk," *Inquiry*, 34, pp. 311-324.
[43] Zweifel, P. and Manning, W.G. (2000): "Moral Hazard and Consumer Incentives in Health Care," Handbook of Health Economics, Volume 1A, 2000, *Elsevier*, pp. 409 - 459.

Facoltà di Economia, Università di Palermo
E-mail address: vdardano@unipa.it

Department of Economics, University of York and Facoltà di Economia, Università di Palermo
E-mail address: pld500@york.ac.uk