EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 03/12

February 2012

Incentive and Selection Effects of Medigap Insurance on Inpatient Care

by

Valentino Dardanoni

(Università di Palermo)

Paolo Li Donni

(Università di Palermo)

INCENTIVE AND SELECTION EFFECTS OF MEDIGAP INSURANCE ON INPATIENT CARE

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 inpatient care, separating the *incentive* and *selection* effects of supplementary insurance. For this purpose, we use two alternative estimation methods: a standard recursive bivariate probit and a discrete multivariate finite mixture model. We find that estimated incentive effects are modest and quite similar across models. On the other hand, there seems to be very significant selection when one conditions only on variables used by Medigap insurers, with the presence of both adversely and advantageously selected individuals, stemming from the multidimensional nature of residual heterogeneity.

JEL Classification Numbers C52, D82, G22, I10.

Keywords Health Insurance; Asymmetric Information; Medigap; Health Care Demand; Finite Mixture Models.

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 the presence of an incentive effect - usually called *ex-post moral hazard* - is very well known by the theoretical literature on contract theory (Arrow[4], Pauly [50] and Zweifel and Manning [59]), its empirical relevance is still debated in the literature see Abbring *et al.* [1]-[2], Buchmuller *et al.* [7], Cardon and Hendel [11], Cohen [16], Schellhorn [54], and Cohen and Spiegelman [17] for a review. A major difficulty in estimating the presence of

Date: February 8, 2012.
Valentino Dardanoni: Dipartimento SEAF, Università di Palermo, email: valentino.dardanoni@unipa.it; Paolo Li Donni: Dipartimento SEAF, Università di Palermo, email: paolo.lidonni@unipa.it. We thank two anonymous referees for extremely useful comments, which have helped us to considerably improve the paper.

moral hazard in Medigap insurance is the existence of *self-selection*, since individuals who expect high health care costs may choose a more generous coverage and then ex-post purchase more services.

In an important body of literature following the seminal paper by Chiappori and Salanié [15], the presence of asymmetric information in insurance markets is appraised using the so called "positive correlation" (PC) test (see two recent reviews by Cohen and Spiegelman [17] and Einav *et al.* [27]), which rejects the null hypothesis of no asymmetric information when there is a positive correlation between insurance purchases and risk occurrence, conditional on the individual characteristics used by insurers to price contracts. The PC test, however, cannot disentangle incentive and selection effects, since finding a positive insurance coverage-risk occurrence correlation in the data does not provide conclusive evidence whether there is adverse selection into insurance contracts, moral hazard, or both.

As discussed in the review by Cohen and Spiegelman [17], some empirical studies have found evidence of adverse selection in various insurance markets, including automobile, annuities, life, reverse mortgages and crop. Adverse selection has also been found in very recent applications to health insurance such as Olivella and Vera-Hernandez [49]), Carlin and Town [12]), and Lustig [45]. However, contrary to the theoretical prediction of the standard adverse selection insurance model, a number of empirical studies of insurance markets have found *negative* risk-coverage correlation, a phenomenon which has been named *advantageous* selection.[1] In a seminal paper, Finkelstein and McGarry [31] study the US long-term care insurance market and find negative correlation between insurance purchase and nursing home use, and argue this is due to the existence of two conflicting sources of private information, namely individuals' actual risk and risk attitudes. Negative risk-coverage correlation has also been found in a recent paper by Fang, Keane and Silverman [30], who show that individuals with Medigap insurance tend to spend *less* on medical care. This is explained again by the existence of multiple dimensions of private information, with cognitive ability being one of the key sources of advantageous selection.[2]

There are different ways to distinguish empirically selection from incentive effects. A strategy is to use experimental data such as the RAND Health Insurance Experiment (RHIE), where to identify the incentive effect controlling for 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.* [13] and Eichner [24]). Einav, Finkelstein and Cullen [26] recently lay down a novel framework for analyzing private information in insurance markets by means of a model which estimates the demand and cost

[1]Alternatively, some authors refer to this situation as *favorable* or *propitious* selection. See Hemenway [36] for an early discussion of the relevance of advantageous selection in insurance markets.

[2]The possibility that multidimensional private information may invalidate standard insurance model predictions has also been the subject of recent theoretical work (see e.g. Chiappori *et al.* [14], de Meza and Webb [20], Smart [55], Villeneuve [56], Wambach [57]).

curves for the insurance contract. This framework allows not only a simple test for the presence and nature of selection effects based on the estimated cost curves (MC test), but also provides a tool to evaluate the policy implications of private information. While the Einav and Finkelstein model is an excellent example of the integration of economic theory and empirical investigation, it requires access to insurance administrative data with information on premiums and claim-level expenditures, and, crucially, *exogenous* identifying variation in prices.

In non-experimental settings most studies use large observational data sets which include information on health care use and insurance status. Researchers typically either estimate health care utilization with a probit (Hurd and McGarry [40]) or two-parts (Ettner [29], Khandker and McCormack [44]) model using health indicators to mitigate the presence of unobserved heterogeneity in health status, or estimate insurance coverage and claims considering selection on observable and unobservable factors, by means of a bivariate probit model with a recursive structure between insurance and health care utilization (Holly, *et al.* [37], Jones *et al.* [41], Buchmuller *et al.* [7]).

The aim of this paper is to estimate incentive and selection effects of Medigap insurance on inpatient care, by employing two different empirical strategies. We first lay down a simple theoretical framework for interpreting selection and incentive effects with observational data on claims and insurance purchase. We then discuss the use of a standard recursive bivariate probit for estimating incentive and selection effect, and in alternative we set a multivariate discrete finite mixture model which may be more appropriate for detecting selection effects. We use data from the Health and Retirement Survey (HRS), which contains information on a rich set of variables concerning health status and individual preferences for risk, and use Medigap purchase and hospital admissions as binary dependent variables representing insurance purchase and health care utilization.

The paper is organized as follows. In section 2 and 3 we model and estimate the incentive and selection effects. Section 4 describes the data. Sections 5 and 6 report estimation results and some concluding remarks respectively.

2. Modeling incentive and selection effects

Suppose we observe a binary variable $I \in \{0, 1\}$ which takes value 1 if an individual has bought a supplemental medical insurance contract, and a binary variable $L \in \{0, 1\}$ which takes value 1 if the individual incurs a given medical care loss. In general, I and L need not be binary variables, but frequently the researcher can only observe whether the individual occurred in the risk or whether she is covered by an insurance plan. In our application, L denotes having used hospital care in a given time period, and I having bought Medigap insurance.[3]

[3]Appendix C provides more details on the Medicare and Medigap health insurance system.

Standard economic theory predicts that risk occurrence and insurance coverage are positively correlated and this relationship depends on two sources (Rothschild and Stiglitz [53] and Arnott and Stiglitz [3]). On the one hand when individuals have private information about their actual risk, the insurance contract will be *adversely selected*, with high risk individuals choosing higher insurance coverage. On the other hand, the insurance contract may give the *incentive* to increase risk occurrence by increasing the probability to incur in the risk (ex-ante moral hazard) or by increasing utilization (ex post moral hazard). Both sources of asymmetric information (adverse selection and moral hazard) will cause the same observed positive correlation in the data. The predicted positive correlation between risk and coverage has inspired the seminal contribution by Chiappori and Salanié [15], who considered the testable implications of asymmetric information in insurance markets, and proposed the so called Positive Correlation (PC) test. The PC test rejects the null of absence of private information in a given insurance market when, conditional on consumers' characteristics used by insurance companies to price contracts, individuals with more coverage experience more of the insured risk.

Let $\boldsymbol{x}$ denote the set of variables used by insurance companies to price a given insurance contract. Conditioning on $\boldsymbol{x}$ is crucial to properly account for asymmetric information, since without conditioning on $\boldsymbol{x}$ it would not be possible to know whether a correlation arises because individuals which are offered the same contract have different risk or rather because they face contracts at different prices (Chiappori and Salanié [15], Einav and Finkelstein [25]). In practice, in many insurance markets there are observable characteristics which *cannot* be used either by regulatory laws or political economy concerns. For example as mentioned in Appendix C, in the Medigap market insurance companies can price contracts mostly on the applicant's age and gender. Thus, a natural choice is to use as conditioning variables those which are effectively used by insurers to price contracts so as to look at how people behave conditional on the menu of contracts they actually face, and in this sense unobserved heterogeneity is implicitly captured by whatever affects L and I but is not used by insurance companies to price contracts.

To clarify the nature of incentive and selection effects in this contests, let $R \in \{0, 1\}$ denote a private information variable describing risk types, so that $R = 0$ denotes low risks and $R = 1$ denotes high risks. The question is how the joint distribution of the observables (L, I), conditional on $\boldsymbol{x}$, is affected by incentive and selection effects in the presence of the unobservable R.

2.1. **Incentive and selection effects.** For simplicity, assume initially that the analysis is conditional on a given value of the observables $\boldsymbol{x}$ used by the insurance company to price contracts.[4]

[4]In this section when not explicitly stated all probabilities should be interpreted as conditional on a given value of $\boldsymbol{x}$, which we omit for simplicity.

To focus first on selection, let us initially assume that there is no moral hazard:

Assumption 1. $P(L=1 \mid R, I=0) = P(L=1 \mid R, I=1)$.

The classic RS definition of adverse selection, that higher risk individuals (i.e. those with higher loss probability) are more likely to buy insurance, implies that insurance purchase and loss probabilities are *comonotone*:

Definition 1. (Adverse Selection 1)*: The insurance contract I is adversely selected if*

$$\big(P(L=1 \mid R=1) - P(L=1 \mid R=0)\big) \cdot \big(P(I=1 \mid R=1) - P(I=1 \mid R=0)\big) > 0.$$

An alternative definition of adverse selection in this context can be derived by comparing the loss and insurance purchase probabilities for the two risk types to the *average* purchase and loss and purchase probabilities, denoted $\bar{P}_L$ and $\bar{P}_I$;[5] the insurance contract is adversely selected by type $R = i \in \{0, 1\}$ if

$$\big(P(L=1 \mid R=i) - \bar{P}_L\big) \cdot \big(P(I=1 \mid R=i) - \bar{P}_I\big) > 0.$$

Definition 2. (Adverse Selection 2)*: The insurance contract I is adversely selected if it is adversely selected by types $R = 0, 1$.*

Definitions 1 and 2 are not directly testable since they involve the unobserved variable R. Consider then the following two testable conditions:

Definition 3. (Positive Correlation Test 1)*:*

$$P(L=1 \mid I=1) > P(L=1 \mid I=0),$$

and

Definition 4. (Positive Correlation Test 2)*:*

$$\frac{P(L=1, I=1) \cdot P(L=0, I=0)}{P(L=1, I=0) \cdot P(L=0, I=1)} > 1.$$

Definitions 3 and 4 are two ways to implement the PC test in this context. Definition 3 says that the expected loss for consumers who chose to insure is greater than for consumers who did not. Definition 4 says that the odds ratio between L and I should be greater than one, that is, L and I should be positively correlated, and goes back to the original implementation of Chiappori and Salanié [15].[6] The following result, which is proven in Appendix A, clarifies the relationship between the four definitions:

Proposition 1. *Definitions 1–4 are equivalent under Assumption 1.*

[5]That is, $\bar{P}_L = P(R=0)P(L=1 \mid R=0) + P(R=1)P(L=1 \mid R=1)$ and $\bar{P}_I = P(R=0)P(I=1 \mid R=0) + P(R=1)P(I=1 \mid R=1)$.

[6]Both alternative definitions of the PC test are discussed in the literature; for example, the review of Einav, Finkelstein and Levin [27] privileges the first, while the review of Cohen and Spiegelman [17] considers both.

The proposition shows that without incentive effect, adverse selection, which involves the unobservable R, is equivalent to the positive correlation property which involves only observables. For *advantageous* selection one simply reverses the inequalities above.

How does the presence of moral hazard affect testing for selection effects? Under moral hazard, insured individuals have a higher expected cost *conditional on their risk type*, so we can model (I, L, R) as

$$\begin{array}{l} P(I=1 \mid R=r) = F\big(a_I(r)\big),\ r \in \{0,1\}, \\ P(L=1 \mid R=r, I=i) = F\big(a_L(r)+bI\big),\ r \in \{0,1\},\ i \in \{0,1\} \end{array} \quad (1)$$

where F denotes a chosen link function, such as probit or logit. In words, b represents the ex-post increase in expected costs (in the chosen probit or logit scale) which is incurred after individuals have selected the insurance contract. In particular, the incentive effect captured by b is *net* of the moral hazard induced by the selection of a contract by different types, who anticipate that their behavior will change after buying the contract (e.g. they adapt to the insurance coverage structure).[7] This anticipated effect is naturally included in the selection effect and is incorporated in the type-specific constants $a(r)$, while the incentive effect b can be interpreted as the residual unanticipated effect of buying insurance on the loss occurrence.

Under moral hazard (i.e. when Assumption 1 does not hold), the equivalence of the PC testing procedures (Definitions 3-4) and the selection definitions (Definitions 1-2) breaks down. How do these definitions change in model (1) which incorporates moral hazard? In Definition 1 of adverse selection, since we are interested in the expected cost of the individuals who buy the contract, replace the marginal loss probability with the insureds' loss probability:

$$\big(P(L=1 \mid R=r, I=1) - P(L=1 \mid R=0, I=1)\big) \cdot \big(P(I=1 \mid R=1) - P(I=1 \mid R=0)\big) > 0.$$

It thus follows that under model (1), which takes into account the presence of moral hazard, the contract is adversely (advantageously) selected when:

$$\big(a_L(R=1) - a_L(R=0)\big) \cdot \big(a_I(R=1) - a_I(R=0)\big) > (<) 0. \quad (2)$$

In practice, rewrite model (1) (including now for completeness the pricing variables $\boldsymbol{x}$) as

$$\begin{array}{rcl} I &=& 1(a_I(R=r) + \boldsymbol{x}'\boldsymbol{\beta}_I + \xi_I > 0), \\ L &=& 1(a_L(R=r) + \boldsymbol{x}'\boldsymbol{\beta}_L + bI + \xi_L > 0) \end{array} \quad (3)$$

[7] Recent studies refer to this effect as *selection on moral hazard* and interpret it as the possibility that individuals select into insurance coverage by anticipating their behavioral response to the coverage structure (see Einav *et al.* [28] who showed that selection on moral hazard can be an empirically relevant component of moral hazard.)

with 1(.) denoting the indicator function, and ξ_I and ξ_L being idiosyncratic errors with continuous c.d.f F. In model (3), $a_I(r)$ and $a_L(r)$ can be used to evaluate selection according to (2), while b estimates the incentive effect.

3. The estimation of the incentive and selection effects

3.1. The recursive bivariate probit model.

A standard approach to estimate model (3) is by using the recursive binary probit model

$$\begin{aligned} I &= 1(\boldsymbol{x}'\boldsymbol{\beta}_I + \epsilon_I > 0), \\ L &= 1(\boldsymbol{x}'\boldsymbol{\beta}_L + bI + \epsilon_L > 0) \end{aligned} \tag{4}$$

with ϵ_I and ϵ_L being standard normal errors with correlation ρ, which captures unobserved selection. Comparing models (3) and (4), we see that in the standard recursive bivariate probit the correlation in ϵ_I and ϵ_L depends on the underlying behavior of the unobserved types, so that the null of the absence of selection amounts to testing $\rho = 0$, with a positive (negative) ρ indicating adverse (advantageous) selection. The null of no incentive effect can be tested by imposing $b = 0$.

In our application, I is a binary variable indicating purchase of supplementary Medigap insurance in 2006 and L is a binary variable indication use of hospital services in 2006. Since Medigap insurance pricing is almost entirely based only on the applicant's age and gender, in our application $\boldsymbol{x}$ denotes gender and a third degree polynomial in age, with interactions.

3.1.1. *An extended recursive bivariate probit model.*

Estimating (4) with a recursive bivariate probit to disentangle incentive and selection effects has some theoretical and econometric difficulties. From an econometric point of view, identification of model (4) relying on the normality assumption, though theoretically feasible as long as data on $\boldsymbol{x}$ are of full rank (see Wilde [58]), it is quite fragile in the absence of exclusion restrictions. A common strategy to check the robustness of the recursive bivariate probit estimates is to compare estimates to those obtained by performing separate probit regressions for the two response variables. It is not uncommon, however, that these differ substantially, even in the case when a standard LR test does not reject the null of $\rho = 0$. When this happens, it can be taken as evidence of potential identification problems in the bivariate probit model.

In addition to these econometric issues, from the theoretical perspective the bivariate normality assumption can be severely limiting when residual heterogeneity is actually multidimensional. Following Finkelstein and McGarry [31] suppose for example there are two conflicting sources of private information, say individual's actual risk R and risk aversion A. For a simple example, suppose the probit errors can be written as $\epsilon_I = A + \eta_I$ and $\epsilon_L = R + \eta_L$, with η_I, η_L uncorrelated idiosyncratic errors, so that risk aversion affects insurance purchase and risk type affects actual loss. If R and A are independent, there is zero residual correlation between I and L ($\rho = 0$), which is interpreted as evidence of no selection. However, the contract is, for example, adversely selected by individuals with high risk aversion A and high risk R since, compared to the population, they are *both* more likely to buy insurance *and* to have a loss; and at the same time the contract is advantageously selected by individuals with high A and low R since they tend to buy the contract but not to incur the loss.[8]

A natural strategy for better identification of the model is to extend model (4) by adding a vector of additional controls which affect insurance purchase and medical care use but are not used by insurance companies to price contracts (i.e., do not belong to $\boldsymbol{x}$). In practice, regulatory constraint and political economy considerations often restrict quite considerably the set $\boldsymbol{x}$, so there are many observable variables which are good candidates for explaining I and L. Let $\boldsymbol{w}$ denote a set of demographic and socio-economic variables which are typically included in medical care use applications, and $\boldsymbol{z}$ a set of observable *proxies* for the unobservable characteristics such as risk tolerance, attitude towards medical care use and insurance purchase, actual (health) riskiness and so on. Examples of such proxies in this context are cognitive abilities, occupational risk, risk reducing or increasing behavior such as preventive care, seat belt use, smoking and drinking, past insurance choices and claims, and objective and subjective health status indicators. In our Medigap application, $\boldsymbol{w}$ contains education and income; $\boldsymbol{z}$ is made of four binary variables (namely, supplementary insurance purchase and hospital use in 2002 and 2004), and two discrete ordered health status variables.

An extended recursive bivariate probit is then

$$\begin{aligned} I &= 1(\boldsymbol{x}'\boldsymbol{\beta}_I + \boldsymbol{w}'\boldsymbol{\gamma}_I + \boldsymbol{z}'\boldsymbol{\delta}_I + \theta_I > 0), \\ L &= 1(\boldsymbol{x}'\boldsymbol{\beta}_L + \boldsymbol{w}'\boldsymbol{\gamma}_L + \boldsymbol{z}'\boldsymbol{\delta}_L + bI + \theta_L > 0) \end{aligned} \tag{5}$$

where sharper parameters identifiability can be obtained with appropriate inclusion/exclusion restrictions setting to zero appropriate elements of $\boldsymbol{w}$ and/or $\boldsymbol{z}$ in the two equations.[9] Model (5) can be considered as the standard applied econometric tool for studying health care insurance markets (see e.g. Ettner [29], Holly *et al.* [37], Buchmuller *et al.* [7], Jones *et al.* [41]). By appropriate inclusion/exclusion restriction, it can allow a much more robust parameter identification compared to model (4) above. However, it still suffers from some limitations. Compared with the simple model (4) above, the extended model makes detection and interpretation of selection effects problematic, since the residual heterogeneity which affects claims and insurance coverage given current pricing $\boldsymbol{x}$ does not have anymore a clear and direct interpretation.

[8]When there are more than 2 types, the equivalence of the two selection definitions 1 and 2 breaks down; to be precise, here we discuss *local* selection by type by comparing claims and coverages to the population averages as in Definition 2.

[9]Using lagged variables to proxy unobserved heterogeneity is a common strategy in this framework (see e.g. Abbring *et al.* [2], Jones *et al.* [41]). It is well known that lagged dependent variables may affect the consistency of parameter estimates. In a very recent paper, de Jong and Woutersen [42] show consistency of probit estimates whenever the error distribution conditional on exogenous covariates and lagged regressors is standard normal. As robustness check, we estimate model (5) without lags to see whether estimated incentive effect b changes significantly.

Summing up, while a carefully set model (5) may in principle be capable of accurate estimation of the incentive effect, it may difficult to use the model for the understanding of selection effects, especially under multidimensional private information. Thus, the objective of the extended model is not to estimate selection, because one is including variables that are not used for pricing; on the other hand, the estimate of the incentive effect in (4) and (5) should be similar if both models properly account for selection.

3.2. **A discrete multivariate finite mixture model.** An alternative procedure to estimate model (3) allowing a clear understanding of selection effects under multidimensional private information is by using a latent class model which identifies a finite number of "types", which differ with respect to their attitude to buy insurance and use medical care.

Suppose there is a set of residual heterogeneity variables which affect insurance choice and outcomes after conditioning on $\boldsymbol{x}$. If we assume that such variables are discrete (a fairly innocuous assumption since any continuous variable can be approximated arbitrarily well by a discrete one), we can cross-classify them into a single discrete variable which identifies the set of heterogeneous "types". Let then U be a random variable with support in $\{1, 2, \ldots, m\}$ which cross-classifies all relevant unobserved heterogeneity. Clearly since the label of the types is arbitrary, no order is assumed on U: for example, with two binary private information variables capturing risk type and risk preference, as shown by Smart [55] there is no ordering of the four types since single crossing of the indifference curves fails to hold. In this setting, different types are simply meant to capture heterogeneous insurance and claim behaviours after conditioning on $\boldsymbol{x}$, without any assumption on the underlying structure; what matters here is that a sufficient number of types is used to capture all heterogeneity.

To analyze selection and incentive effects of the insurance contract we then need to estimate (cfr. model (1))

$$\begin{array}{rcl} I &=& 1(\sum_{u=1}^{m} \alpha_I(u)U(u) + \boldsymbol{x}'\boldsymbol{\beta}_I + \eta_I > 0), \\ L &=& 1(\sum_{u=1}^{m} \alpha_L(u)U(u) + \boldsymbol{x}'\boldsymbol{\beta}_L + bI + \eta_L > 0) \end{array} \tag{6}$$

where $U(u), u = 1, \ldots, m$ denotes a set of m mutually exclusive dummy variables defining types' membership, so that the α coefficients can be interpreted as random intercepts with a nonparametric discrete specification, like in Heckman and Singer [35]. Since no underlying structure is imposed on U, the α's can capture in a unrestrictive way any variable which affects medical care use and insurance demand after conditioning on $\boldsymbol{x}$, and thus take into account the potential multidimensionality of residual heterogeneity. (η_I, η_L) can be seen as idiosyncratic errors, which are uncorrelated since, in the spirit of finite mixture models, U makes I and L conditional independent. In this sense, comparing (6) and (4), the finite mixture model decomposes residual heterogeneity by "uncovering" the deep types which systematically affect claims and coverage.

Model (6) is completed by the types membership probabilities. To force the types probabilities to lie between zero and one and sum to one, it is convenient use a multinomial logit parameterization:

$$P(U=u)=\frac{\exp(\alpha_U(u))}{\sum_{u=1}^{m}\exp(\alpha_U(u))},\ \alpha_U(m)=0,\ u=1\ldots,m. \tag{7}$$

Notice that the $m-1$ logit parameters α_T are simply reparametrizations of the membership probabilities, and do not impose any parametric restriction on the distribution of U.

Within model (6)-(7) there is an incentive effect when $b>0$. Since the labeling of the types is arbitrary, selection effects can be identified by first relabeling the types according to their willingness to buy insurance $\alpha_I(1)\geq\cdots\geq\alpha_I(m)$, and then if $\alpha_L(1)\geq\cdots\geq\alpha_L(1)$ the market is adversely selected, while if $\alpha_L(1)\leq\cdots\leq\alpha_L(1)$ there is advantageous selection. If, on the other hand, the α_I's and α_L's are not comonotone, then locally there are types which adversely/advantageously select the contract.[10] Model (6)-(7) can be seen as an instance of a *semiparametric finite mixture model*, where the parameters α_I's and α_L's can be interpreted as random intercepts with a nonparametric discrete specification, like in Heckman and Singer [35], and the parametric part is given by the choice of the errors' distribution in (6) (we use the standard logistic distribution). Well known applications of this framework are Deb and Trivedi ([21] and [22]) in health economics and Cameron and Heckman [10] in education.

The semiparametric finite mixture model is a well known and well established tool to capture residual heterogeneity. However, in our bivariate binary context, it could be limited in the identification of multiple unobserved types. In fact, conditionally on $\boldsymbol{x}$, we observe only two binary variables (that is, 3 parameters), while even with only two mixture types there are 5 parameters to estimate. Thus, it must rely on covariates' variation to achieve parameter identification, which in our Medigap application are limited to age and gender. In practice, model (6)-(7) typically achieves identification of very few unobserved types; in many applications only two types are identified.

To achieve shaper identification of residual heterogeneity, in spirit with Finkelstein and Poterba [32], Finkelstein and McGarry [31] and Cutler, Finkelstein and McGarry [19], we may obtain identifying variation by using a set of observable variables $\boldsymbol{z}$ unused by insurers for pricing but acting as *indicators* of U (that is, observable manifestations of the unobserved heterogeneity which affects insurance choice and medical care use). Thus, we may set an *auxiliary system of equations*, which is estimated jointly with the main equations (6) and the membership probability equation (7) to help identification of the heterogeneous types.

The econometric model we use is based on the decomposition of a joint multivariate distribution of a set of discrete variables into a finite mixture of conditionally independent

[10] This testing procedure uses the same motivation of Einav *et al.* [26] MC test, since both tests are based on the comonotonicity of willingness to pay and marginal cost. The MC test rejects the hypothesis of no selection if the MC curve is constant, whereas the selection is adverse (advantageous) if the MC curve is increasing (decreasing) in price.

univariate distributions, and thus requires that these indicators are *discrete ordered variables*.[11] Suppose we have k auxiliary variables, with $z_j \in \{0, \dots, l_j\}$, say, for $j = 1, \dots, k$. In our application, we use past insurance choice and hospital use, who are actually binary variables (so that l_j is actually equal to 1), and health status indicators which take 5 ordered values. Appendix B discusses how the auxiliary system may help identification of latent parameters by increasing the number of observable ones; the key intuition is that adding observable responses z_j's increases observable parameters at a much slower rate that latent ones.

The auxiliary system of equations can then be written as a standard ordered regression model

$$z_j^* = \sum_{u=1}^{m} \alpha_{z_j}(u)U(u) + \epsilon_j, \; j = 1, \dots, k \tag{8}$$

where $z_j = l \leftrightarrow \delta_{j,l-1} \leq z_j^* \leq \delta_{j,l}$, for $l = 1, \dots l_j$ and $j = 1, \dots, k$, and $\epsilon_1, \dots, \epsilon_k$ are independent idiosyncratic standard logistic errors. When z_j is binary, equation (8) estimates m conditional probabilities, while when z_j takes more than two values there are m random heterogeneity intercepts and $l_j - 1$ cut points.

The discrete multivariate finite mixture model is defined by the main equations (6), the membership probability equation (7), and the auxiliary equations (8). It can be seen as an instance of a discrete multivariate MIMIC (Joreskog and Goldberger[43]) model (see Goodman [34] for the seminal paper on finite mixture models with multivariate binary responses, and Huang and Bandeen-Roche [39] for a recent general treatment). Differently to the MIMIC model, the residual heterogeneity U is not a continuous univariate variable on the real line, but an unstructured nonparametric variable.

3.3. **Variations in the model.** For the purposes of detecting selection and incentive effects, in model (6-7-8) we are mainly interested in estimating the parameters of the two main equations 6. However, alternative specifications of the membership probabilities in equation (7) and of the auxiliary equations in equation (8), by identifying different sets of heterogenous types U, may of course influence parameter estimates in the main equation (6). For example, controlling for education in the auxiliary and in the membership probability equations helps identifying types according to their education-adjusted (i.e. relative) health condition, propensity to insure, risk tolerance etc. Thus, for robustness, it is important to expand equations 7 and 8 to take into account further appropriate controls.

In our application, after estimating the simplest model (6-7-8), we also estimate a more complete model where: i) the multinomial membership probabilities and all the indicators z_j depend on $\boldsymbol{x}$ and on demographic $\boldsymbol{w}$; ii) the two past hospital use indicators in addition depend also on insurance purchase. Additional models which selectively add controls to different components of 7 and 8 may be also estimated for robustness.

[11]There are various different models with a discrete latent variable and continuous indicators which impose different distributional assumptions, see e.g. Bartholomew and Knott [5]

3.4. **Estimation of the finite mixture models.** Estimation of the multivariate finite mixture models can be obtained by the EM algorithm, which is the standard approach for maximum likelihood estimation of finite mixture models, and has been shown (Dempster *et al.* [23]) to converge to the maximum of the true likelihood. A detailed discussion of the EM algorithm in a system of non linear structural equations with latent classes can be found in Bergsma *et al.* [6].[12]

It is well known that the EM algorithm may converge even if the model is not identified, which is a crucial issue for finite mixture models. Conditions for parametric identification are discussed in Theorem 1 of Huang and Bandeen Roche [39]; local identification can be checked using the the numerical test described by Forcina [33], which consists in checking that the Jacobian of the transformation between the parameters of the observable responses and the mixture model parameters is of full rank for a wide range of parameter values. Preliminary to estimation, we have run this test for all models considered, and they all seem solidly identified.

Despite the usefulness of finite mixture models to detect underlying residual heterogeneity, one unresolved issue in their application is how to determine the number of unobserved types m. The currently preferred approach suggests the use of Schwartz's Bayesian Information Criterion (BIC) to guide this choice, which in certain conditions is known to be consistent and generally helps preventing overparametrization (see McLachlan and Peel [47] for a thorough introduction to finite mixture models and a review of existing criteria for the choice of the number of types). BIC is calculated from the maximized log-likelihood L by penalizing parameters' proliferation, $BIC = -2L + vlog(n)$, where n denotes sample size and v the number of parameters; the model with the lowest BIC is preferred.

Finally, it may be worth noticing that, since the number of types m is not predetermined, formal hypotheses tests performed in finite mixture models are in fact conditional on m, and pre-testing for the number of types may invalidate distributional results of the test statistics employed. Of course, while pre-testing is a common problem in most applied research whenever final estimates are obtained after searching for appropriate specification, this is an issue which should be kept in mind whenever test results differ significantly when performed under different values of m.

4. Data and 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. We use

[12]We are grateful to Antonio Forcina for kindly providing the Matlab code for the EM estimation. Estimation has been done with a code rewritten in STATA, which is available upon request.

the 2006 wave as reference point and collect information from the previous two waves (2002 and 2004). Since we study the effect of supplemental insurance (Medigap) on health care in the Medicare system (see Appendix C for a detailed discussion of this insurance market), we also exclude those individuals that received additional coverage through a former employer, spouse or some other government agency. Following Fang *et al.* [30] we define an individual as having additional health insurance coverage (Medigap) if she purchased directly health insurance policy in addition to Medicare.

Since our focus is to disentangle incentive and selection effects at time 2006 for those individual who deliberately choose to purchase additional coverage, our sample should be limited only to people who are covered by Medicare, are not covered by additional public insurance and pay personally the required monthly premium. Therefore we exclude those who are enrolled in any other public program different from Medicare or that are covered at year 2006 by Medigap insurance plan provided by own or spouse's former employer. For this reason we consider only individuals older than 65 in 2002.

Our sample is composed by 3368 individuals and descriptive statistics are reported in table 4. Supplemental insurance is coded as a binary variable which takes value 1 if respondent has any (no long-term care) supplemental private Medigap insurance coverage. Almost 42 percent of Medicare beneficiaries in the sample has a supplemental insurance, and 87 percent of them are continuously covered since they turn 65; thus, most of Medicare beneficiaries, if they purchase a supplemental insurance coverage they do it as soon as they are enrolled in Medicare. In addition to these variables we also use information on whether additional coverage was provided in the previous years 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, and can be considered as appropriate instruments since they are likely to affect insurance choice but they can be excluded from the inpatient care equation (Ettner [29] finds that selection into additional coverage can be driven by employer-provided plans since high-risk (or risk averse) workers may self-select into jobs that provide better retirement medical benefits).

The HRS offers detailed information on health care consumption. We focus on hospital staying over the three waves (h02, h04 and h06). Hospital admissions account for 29% of the Medicare's total expenditure and it is an important part of the Medicare total expenditure (see CMS, Office of the Actuary, National Health Expenditure Accounts, 2007). These variables are binary and take value 1 if individual had at least one hospital admission, and 0 otherwise. As explained in details in Appendix C, in the Medigap market insurance companies are constrained by Federal law to use only age and gender to price contracts. Therefore as variables used by insurer to predict risk we only use these two measures; specifically we set as controls a polynomial in age of order three with each term multiplied by gender to allow for differential pricing of women and man.

We follow previous studies on health care demand (Cameron *et al.* [9], Jones *et al.* [41], Deb and Trivedi [22]) and insurance choice (Propper [51], Cameron and Trivedi [8]) as guidance for selecting relevant controls. Health conditions have an important influence both on the decision to subscribe supplementary insurance as well as on the utilization of health care services. We include in the analysis the number of self-reported doctor diagnosed longstanding or chronical diseases (DIS) - such as high blood pressure, hypertension, cardiovascular disease, lung disease, kidney conditions, emotional and psychiatric problems - and the index of Activities of Daily Living (ADL), which measures difficulties in bathing, dressing, eating, getting in/out of bed and walking across a room. Both indices are defined on a discrete scale ranging between 0-4, and are averaged out over the two previous years period to capture the persistency of need status and any preexisting condition which may affect insurance choice over the time. The last group of variables includes socio-economic characteristics, such as education and wealth. Education is measured using three dummy variables, while individual wealth is measured as a continuous variables averaging individual wealth over the three waves. As reported in Table 4, we find that people with additional insurance tend to have higher education, to be in the top wealth quartile, and to have a lower score of ADL, though the average number of diseases does not vary substantially.

5. Results

5.1. **Single probit and bivariate probit results.** We first estimated a bivariate probit model using only the age and gender pricing variables as controls. As mentioned in section 3.1, parameter identification is a potential issue in the absence of inclusion/exclusion restrictions. This is evidenced by the conflicting results of the Wald test for the null $\rho = 0$ (which is rejected with p-value 0.018) and the likelihood ratio test which does not reject the null with p-value 0.792. As argued by Monfardini and Radice [48], the LR test in general performs significantly better compared to the Wald test, but when the two tests give contrasting results this usually signals identification problems.

We then turn to the extended model in equation (5), which includes additional controls, which help to identify the incentive effect and act as proxy the underlying unobserved heterogeneity. In the extended model, we add in both equations income, education, the two measures of health status, and lagged values of hospital use and insurance purchase. For identification, we also exclude some variables from the outcome (hospital) equation which are correlated with insurance choice but, conditional on exogenous variables, uncorrelated with hospital utilization. Following Ettner [29], in our specification we have adopted as instruments past supplemental insurance coverage provided by own (iemp04) or spouse's (iemps04) former employer.

Table 6 in Appendix D shows the single equation probits and the recursive bivariate probit estimated coefficients of insurance choice and inpatient stay for the extended model (5). Estimated parameters from the recursive bivariate probit model reveal that the probability of a hospital staying increases with age and education, but not with wealth, and it is positively associated with low persistent health status and past inpatient stay. On the other hand, the probability of enrolling in a supplementary insurance plan significantly increases with past insurance status, but not with wealth and education. In addition, having better coverage in the past (provided either individually, spins02 and spins04, or by a former employer, iemp04 and iemps04) significantly increases the probability of having supplementary insurance. These results seem to indicate that individuals with a persistent coverage over the years tend to be more likely to purchase supplemental insurance.

The estimated ρ is equal to -0.232, with a s.e. equal to 0.213. The Wald test statistic and LR test statistics are equal to 1.09 and 0.35 and are asymptotically distributed as χ^2_1. Both statistics cannot reject the null $\rho = 0$ with a p-values of 0.294 and 0.555 respectively. Thus, the two testing procedures in this case give very similar conclusions.

However, as argued in section 3.1.1, the extended model does not allow a direct interpretation in terms of selection effects, because of the inclusion of potentially observable heterogeneity that cannot be priced by the insurance companies (due to regulation); thus, the evidence of the absence of residual claims/coverage correlation is not conclusive evidence of absence of selection.

Finally, to check the sensitivity of the estimated coefficients to the assumed exclusion restrictions, we estimated the bivariate probit model under different specifications of the outcome equation, analogously to Buchmuller *et al.* [7]. Table 5 in Appendix D reports results on the two parameters of interest, b and ρ, for several versions of the model. Although the point estimates vary slightly depending on which variables are excluded from the utilization equation, the qualitative results are similar across the different specifications.

A rather puzzling finding is that for all versions of the extended recursive bivariate probit model the estimated incentive effect is much larger (but with a large standard error) than the one from the single equation probit model, even if ρ is never statistically significant. It is plausible to conclude that the errors in the two estimated equations are independent, and the single equation probit model gives a more reliable estimate of the incentive effect parameter b. The average marginal effect in the single equation probit is equal to 0.041.[13]

5.2. **Discrete multivariate finite mixture model.** We now consider model (6-7-8) using different specifications for the type membership probabilities and the auxiliary system as discussed in section 3.3. We initially estimated a large number of different configurations of controls. Interestingly, the calculated BIC indicates that in all the models we considered, four types are adequate to represent residual heterogeneity.

[13]As discussed in footnote 8, for robustness we also estimated model (5) without lags. In the single equation probit case, without lagged variables the estimated incentive effect b is equal to 0.13 with a s.e. of 0.04; in the bivariate probit model the estimated b is equal to 0.40 with a s.e. of 1.5. A comparison with Table 6 reveals that lagged dependent variables do not significantly change the general pattern of estimated incentive effects.

For reasons of space, we discuss only estimation of the "smallest" and "biggest" models, and other two models with the lowest BIC values:

- Model 1 is the smallest model as described in equations (6-7-8) with no controls in the auxiliary and membership probability equations.
- Model 2 is the largest model, which adds age and gender (with interactions), wealth, and education in all auxiliary equations and in the multinomial membership probability equation, and in addition adds supplementary insurance in the hospital use equations h02 and h04.
- Model 3 adds, in model 1, supplementary insurance in the hospital use equations h02 and h04.
- Model 4 adds, in model 1, income and education to the membership probabilities.

Table 1 reports the the maximized log-likelihood L, the Schwartz's Bayesian Information Criterion BIC and the number of parameters in the four models for values of $m = 2, \ldots, 5$.

Table 1. Model Selection Criteria for alternative model specifications

Model 1				
	Number of Latent Classes			
	m=2	m=3	m=4	m=5
L	-19169.19	-18833.72	-18662.63	-18638.94
BIC	38647.02	38049.17	37780.09	37805.81
# of parameters	38	47	56	65
Model 2				
	Number of Latent Classes			
	m=2	m=3	m=4	m=5
L	-18920.65	-18602.25	-18440.83	-18417.65
BIC	38791.58	38317.23	38156.83	38272.9
# of parameters	117	137	157	177
Model 3				
	Number of Latent Classes			
	m=2	m=3	m=4	m=5
L	-19166.43	-18833.01	-18658.88	-18657.07
BIC	38657.75	38064.01	37788.84	37858.31
# of parameters	40	49	58	67
Model 4				
	Number of Latent Classes			
	m=2	m=3	m=4	m=5
L	-19990.16	-18781.05	-18589.38	-18560.93
BIC	40321.45	38008.82	37731.07	37779.74
# of parameters	42	55	68	81

Since, as argued above, in all configurations BIC indicates that four types are adequate to represent residual heterogeneity, we will comment on incentive and selection effects focusing on the case $m = 4$. The results with different number of types (which are not further discussed, but are available upon request) show that the incentive effect parameter b of main interest is quite robust to the number of types $m = 3, 4, 5$, and in all cases there is substantial residual heterogeneity as described by the α coefficients, which seem to vary very significantly across types.

Table 7 in Appendix D contains parameter estimates for model 4, which is the model with the lowest value of BIC. In the sample about 33% is of type one, and about 26%, 17% and 22% are of the remaining three types. Let us first focus on the incentive effect. The parameter b is positive and highly significant. The average marginal effect is equal to 0.045, which is slightly greater than the average marginal effect calculated in the single equation probit model reported in section 5.1 (namely, 0.037 in the simple model and 0.041 in the extended model). The similarity of estimated marginal effects in the probit and the finite mixture models may seem surprising since the former does not account satisfactorily for selection.

The explanation for this puzzle may become clear after the discussion below of the estimated residual heterogeneity, which shows the existence of two unobserved variables which are cross-classified into four types which pull observed I and L in different directions. This may imply that, conditional on $\boldsymbol{x}$, there is not a great deal of marginal association in the joint distribution of I and L (4).[14]

Regarding selection effects, rather than looking at the estimated α parameters, it is more instructive to derive the conditional probabilities of all response variables.[15] Table 2 shows very substantial heterogeneity in hospital admission and Medigap purchase by the four types: types 3-4 tend to have a much higher probability of hospital admission than types 1-2, while types 1-4 have a much lower probability of buying Medigap than types 2-3.[16] Types 1-2

Table 2. Estimated conditional probabilities

	U=1	U=2	U=3	U=4
h06	0.1830	0.1505	0.5566	0.5278
h04	0.1073	0.1279	0.5436	0.5353
h02	0.1077	0.1485	0.5018	0.4634
spins06	0.1379	0.7868	0.7531	0.1085
spins04	0.0679	0.9338	0.9010	0.0887
spins02	0.0917	0.8210	0.8467	0.1357
adl≥1	0.0721	0.0788	0.3649	0.4235
adl≥2	0.0178	0.0195	0.1181	0.1461
adl≥3	0.0063	0.0070	0.0450	0.0568
adl≥4	0.0027	0.0030	0.0198	0.0251
dis≥1	0.8520	0.8836	0.9816	0.9839
dis≥2	0.5018	0.5703	0.9032	0.9142
dis≥3	0.1713	0.2141	0.6571	0.6863
dis≥4	0.0432	0.0561	0.2947	0.3231

are also more likely to have a lower ADL score and number of diseases than types 3-4.

[14]Estimated coefficients imply that the marginal correlation between the unobserved heterogeneity in the two equations is about -.05.

[15]The estimated conditional probabilities are obtained by substituting the estimated parameters reported in Table 7 in the inverse-logit link function and averaging out for covariates.

[16]While it is apparent that residual heterogeneity coefficients α's (and thus the conditional probabilities above) vary substantially across types, we can still check whether this simply reflects sampling variation by testing the equality constraint $\alpha_{h06}(1) = \cdots = \alpha_{h06}(4)$ and $\alpha_{spins06}(1) = \cdots = \alpha_{spins06}(4)$. The LR test statistic rejects the null of no selection with a p-value lower than 10^{-4}.

Thus, it appears that there are at least two different sources of residual heterogeneity: the attitude to buy insurance (which is low for types one and four and high for types two and three), and the propensity to use medical care (which is high for types three and four and low for types one and two). The overall picture which emerges from our estimates is strongly suggestive of multidimensional residual private information, and there appears to be *both* advantageous and adverse selection of Medigap insurance the four types, since the contract is adversely selected by individuals who have the same (high or low) propensity for insurance purchase and hospital admission, and advantageously selected by individuals who have opposite attitudes.[17] Since, conditional on age and gender, all types actually face the same pricing, this suggests the existence of substantial cross-subsidization of high risks at the expense of low ones. In fact, high risks have almost twice the probability of using inpatient care then low risks.

Regarding other parameters, similarly to the probit case, after controlling for residual heterogeneity it seems that age and gender do not significantly impact insurance choice and hospital utilization. On the other hand, education and income seem to affect type's membership probabilities: taking types 1 (who are those individuals with low attitude to medical care use and insurance purchase) as benchmark, more educated and wealthier individuals tend to have higher insurance preference when they have low medical care use, and lower medical care use when they have low insurance preference.

Finally, for robustness we report the incentive and selection effects for the other 3 models. These are collected in Table 3 below.

[17]As discussed in footnote 8 above, we have appraised *local* selection by type by comparing claims and coverages to the population averages as in Definition 2.

Table 3. Incentive and selection effects for alternative model specifications

Model 1				
	U=1	U=2	U=3	U=4
h06	0.1842	0.1558	0.5846	0.5427
spins06	0.1311	0.7766	0.7628	0.1149
Marginal Effect	0.0395			
Model 2				
	U=1	U=2	U=3	U=4
h06	0.1892	0.1573	0.6131	0.5574
spins06	0.1437	0.7751	0.7468	0.1097
Marginal Effect	0.0409			
Model 3				
	U=1	U=2	U=3	U=4
h06	0.1853	0.1547	0.5889	0.5414
spins06	0.1311	0.7787	0.7604	0.1158
Marginal Effect	0.0399			
Model 4				
	U=1	U=2	U=3	U=4
h06	0.1830	0.1505	0.5566	0.5278
spins06	0.1379	0.7868	0.7531	0.1085
Marginal Effect	0.0463			

Estimated probabilities are averaged out for $\boldsymbol{x}$.

Table 3 shows that the estimated incentive effect does not vary substantially across models. Regarding selection effects, estimated conditional probabilities show a very close pattern of unobserved heterogeneity in the two main equations across models. This confirms the general picture that Medigap is an insurance market with substantial multidimensional residual heterogeneity and heavy cross-subsidization of high risk types at the expense of low risk ones, probably due to regulatory constraints which severely limit insurers' pricing policies.

For additional sensitivity we finally estimated also the standard semiparametric finite mixture model (6-7) under different numbers of latent classes. Since identification relies mainly on covariates' variation, we include in the multinomial membership probabilities $\boldsymbol{z}$ and $\boldsymbol{w}$ to help identification. The BIC for $m = 2, 3, 4$ are equal to 7681.62, 7637.53 and 7664.83, indicating that three classes are adequate. Parameter estimates are reported in Table 8.

The estimated α coefficients for the main system show that types one have a low propensity of medical care utilization but high propensity to buy insurance, while types three have opposite attitudes. Types two have low propensity of using health care and buying insurance. Thus, it appears that while the multivariate model (6-7-8) allows a finer partition of types, model (6-7) tends to lump different types together. A closer look at estimated parameters however reveals that some of the α's and their related s.e. are very large, suggesting possible identification issues. Possible identification problems are also revealed by the estimated incentive effect b, which is much larger than that in model (6-7-8), but is very imprecisely estimated.

6. Final remarks

In the Medigap insurance market there seems to be substantial residual heterogeneity in insurance coverage and inpatient care use after conditioning on variables used by insurers to price the contract. We have employed two alternative econometric approaches to detect the incentive and selection effects of Medigap:

a standard recursive bivariate model, and a discrete multivariate finite mixture model.

Our estimates of the incentive effect seem to be remarkably similar in all well identified models we considered: individuals who buy Medigap seem, on average, to increase their probability of having an inpatient stay of 4 percentage points. The estimated average marginal effect of Medigap on hospital use is slightly smaller than the difference, in the sample, between the probability of hospital admission of individuals who are covered (0.36) and those who are not covered (0.31) by Medigap.

We also find a substantial degree of selection in the market, driven mainly by two residual heterogeneity variables capturing attitude to buy insurance and health care utilization. This paints a general picture that Medigap is an insurance market with substantial multidimensional residual heterogeneity and heavy cross-subsidization of high risk types at the expense of low risk ones. Since the Medigap market is heavily regulated, this suggests further research investigating the welfare effects of regulatory practices which severely limit insurers' pricing policies. The theoretical analyses of Hoy [38], Cracker and Snow [18] and Rothschild [52], and the empirical model described in Einav *et al.* [27] seem the right places to start this investigation.

Appendix A. Proof of Proposition 1

To show that 3 is equivalent to 4, rewrite 4 as $P(L=1, I=1) \cdot P(L=0, I=0) > P(L=1, I=0) \cdot P(L=0, I=1)$, substitute $P(L, I)$ with $P(L \mid I)P(I)$, then use $P(L = 1 \mid I) = 1 - P(L = 0 \mid I)$, and simplify; to show that 1 is equivalent to 2, substitute $\bar{P}_L$ and $\bar{P}_I = P(R=0)P(I=1 \mid R=0) + P(R=1)P(I=1 \mid R=1)$ in 2 and simplify. To show the equivalence between 3 and 1 under Assumption (1), use the Law of Total probability

$$P(L \mid I) = P(R=0 \mid I) \cdot P(L \mid I, R=0) + P(R=1 \mid I) \cdot P(L \mid I, R=1)$$

and (1) to get

$$P(L \mid I) = P(R=0 \mid I) \cdot P(L \mid R=0) + P(R=1 \mid I) \cdot P(L \mid R=1).$$

Thus,

$$P(L=1 \mid I=1) - P(L=1 \mid I=0) = \big(P(R=1 \mid I=1) - P(R=1 \mid I=0)\big) \cdot \big(P(L=1 \mid R=1) - P(L=1 \mid R=0)\big)$$

using $P(R=0 \mid I) = 1 - P(R=1 \mid I)$. By the same argument used above to show the equivalence of 3 and 4, it is easily seen that $P(R=1 \mid I=1) > P(R=1 \mid I=0)$ if and only if $P(I=1 \mid R=1) > P(I=1 \mid R=0)$.

Appendix B. Identification

In this section we discuss the intuition on the identification of the finite mixture model (6-7-8). Suppose we want to identify the insurance and loss probabilities in 6 and the types membership probabilities in 7 conditionally on a given value of $\boldsymbol{x}$. In total there are $m+1$ unobserved parameters for the loss equation, m for the insurance demand equation, and $m-1$ membership probabilities, for a total of $3 \times m$. Since we observe only I and L (the joint distribution of two binary variables with 3 free parameters), unless covariates' variation and parametric restrictions are imposed, it is impossible to identify (6-7) even with $m = 2$.

Our strategy is to look for external sources of variation by means of the set of indicators of the unobservable variable U, $z_j = 1, \ldots, k$, which allow to increase the number of observable parameters; and at the same time to impose appropriate conditional independence restrictions, which are typical of finite mixture models, to reduce the number of parameters to be identified.[18] In our application, in the auxiliary system 8 there are 4 binary variables and 2 ordered variables which take 4 levels. Thus, equation 8 adds $4 \times m$ plus $2 \times m + 6$ unobservable parameters to 6-7, for a total of $9 \times m + 6$ unobservable parameters. However, the joint distribution of the observable responses (i.e. I and L which are of main interest and the set of indicators z_j which are instrumental for identification) has $2^6 \times 4^2 - 1 = 1023$ free parameters. Thus, conditionally on $\boldsymbol{x}$, in principle a very large number of types can be identified without relying on covariates' variation.

It is well known that counting the number of observable parameters is only a necessary condition for identification of model parameters; there are well known pathological examples in the literature on finite mixture models which show that this counting condition is not sufficient. More formally, if we let $\boldsymbol{p}$ denote the vector which arrays the unobservable parameters of the mixture model (6-8-7) (the α's and b), and $\boldsymbol{q}$ the vector which arrays the observable joint distribution of I, L and $z_j = 1, \ldots, k$, $\boldsymbol{q}$ is a known function of $\boldsymbol{p}$, say $\boldsymbol{q} = \boldsymbol{f}(\boldsymbol{p})$. Model (6-7-8) is identified if, for every $\boldsymbol{p}'$ and $\boldsymbol{p}''$,

$$\boldsymbol{f}(\boldsymbol{p}') = \boldsymbol{f}(\boldsymbol{p}'') \rightarrow \boldsymbol{p}' = \boldsymbol{p}'' \tag{9}$$

that is, the model is identified if the mapping between the parameters of the observable responses and the mixture model parameters is one-to-one; in other words, the model is identified if the equation $\boldsymbol{q} = \boldsymbol{f}(\boldsymbol{p})$ has at most one solution for each value of $\boldsymbol{q}$.

Since the state of the mathematical art is such that there are no general conditions for the uniqueness of the solution of nonlinear systems of equation, researchers have then applied the weaker *local* identification condition which requires *local* uniqueness of the solution of the equation $\boldsymbol{q} = \boldsymbol{f}(\boldsymbol{p})$. Local identification of the mixture model (6-7-8) is thus ensured when the Jacobian of the mapping $\boldsymbol{q} = \boldsymbol{f}(\boldsymbol{p})$ is of full rank. In practice, local identification can be checked numerically by using the test described by Forcina [33], which consists in checking that the Jacobian is of full rank for a wide range of randomly drawn parameter values. We performed this test with 10000 replications and found the all models we estimate are well identified.

[18]The use of conditional independence assumptions as a possible strategy for achieving nonparametric identification is discussed, among others, in the survey by Matzkin [46].

Appendix C. Health insurance and access to care for elderly in US

C.1. **Medicare.** Medicare is probably the main source of health insurance for all individuals aged 65 in US and the coverage is near universal (about 97% of the elderly have Medicare).[19] The Medicare program 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 to 20% of expenses.

C.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 the Medigap insurance market is highly regulated by 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 the Medigap market designed a particular mechanism favoring the insured: Medigap insurance companies must offer 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 and gender. In the supplemental health insurance market the most popular Medigap plans are C and F, because they cover major benefits and are less expensive than other plans. For example, plan C offers coverage for skilled-nursing-facility coinsurance, foreign-travel emergencies, deductibles that are required under traditional Medicare and other basic benefits like hospital and outpatient coinsurance. All these plans include Medigap A, which is the basic one,

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

the least expensive and least comprehensive. This plan covers several losses; for example it provides (increasing) coverage for daily medicare copayment per day for hospitals stays; it reimburses the full cost of up to 365 additional hospitals days and the partial cost of other services related to doctor's (outpatient) visits, preventive health screening and outpatient prescription drug.

Appendix D. Tables

Table 4. Sample characteristics and variables definition

Variable	Definition of Binary Variables	Full Sample	No Ins.	With Ins.
Insurance Status				
spins06	1 = enrolled in Medigap at 2006.	0.42	0.00	1.00
spins04	1 = enrolled in Medigap at 2004.	0.45	0.21	0.80
spins02	1 = enrolled in Medigap at 2002.	0.43	0.22	0.72
iemp04	1 = additional coverage from former emp. at 2004.	0.08	0.10	0.05
iemps04	1 = additional coverage from spouse emp. at 2004.	0.04	0.05	0.04
Hospital Admission				
h06	1 = entered a hospital in 2005-2006.	0.33	0.31	0.36
h04	1 = entered a hospital 2003-2004.	0.29	0.29	0.29
h02	1 = entered a hospital in 2001-2002.	0.27	0.26	0.28
Variables Used by insurer to price Medigap plan				
fem	1 = female.	0.61	0.59	0.63
age	individual age at 2006.	78.3	78.1	78.7
ageF	interaction between female and age .	48.1	46.6	50.1
Other Controls unused by insurer				
edu3	1 = if individual is high-school graduate.	0.35	0.33	0.39
edu4	1 = if individual has a degree lower than BA.	0.19	0.17	0.20
edu5	1 = if individual has college degree or greater.	0.16	0.16	0.15
totinc	average individual total wealth.	0.12	0.11	0.14
dis	average number of disease.	2.15	2.16	2.14
adl	average ADL.	0.30	0.33	0.27

No ins. and With ins. denote subsamples without and with Medigap coverage in 2006.

Table 5. Partial results for alternative bivariate probit specifications

Variables excluded in the utilization equation	b	ρ	Likelihood Ratio
None	0.124	-	-
	(0.061)		
iemps04	0.514	-0.223	0.75
	(0.428)	(0.247)	
iemp04	0.421	-0.179	0.41
	(0.462)	(0.266)	
iemp04+iemps04	0.529	-0.231	0.35
	(0.367)	(0.213)	

Note: Robust standard errors are reported in brackets.
The sample size for all models is 3368.

Table 6. Estimated parameter for probit model of equation (5)

	Probit Model		Bivariate Probit Model	
Independent Variables	Hospital 2006	Insurance 2006	Hospital 2006	Insurance 2006
spins06	0.124	.	0.529	.
	(0.0603)	.	(0.368)	.
spins04	0.0580	1.405	-0.136	1.405
	(0.0668)	(0.0634)	(0.169)	(0.0633)
spins02	-0.0413	0.735	-0.134	0.733
	(0.0612)	(0.0593)	(0.104)	(0.0594)
h04	0.552	-0.0106	0.547	-0.00967
	(0.0517)	(0.0599)	(0.0532)	(0.0598)
h02	0.165	0.0272	0.161	0.0325
	(0.0537)	(0.0602)	(0.0536)	(0.0603)
dis	0.167	-0.0204	0.168	-0.0199
	(0.0215)	(0.0234)	(0.0214)	(0.0235)
adl	0.104	-0.0610	0.108	-0.0627
	(0.0335)	(0.0379)	(0.0331)	(0.0378)
fem	-0.169	-0.0245	-0.166	-0.0226
	(0.125)	(0.130)	(0.124)	(0.130)
age	0.181	-0.529	0.221	-0.534
	(0.381)	(0.404)	(0.382)	(0.405)
age^2	-0.122	0.489	-0.160	0.509
	(0.392)	(0.421)	(0.396)	(0.422)
age^3	0.0431	-0.114	0.0519	-0.121
	(0.110)	(0.118)	(0.111)	(0.118)
fem*age	0.296	0.0636	0.301	0.0798
	(0.477)	(0.515)	(0.475)	(0.515)
fem*age^2	0.0304	0.157	0.00120	0.129
	(0.471)	(0.518)	(0.472)	(0.519)
fem*age^3	-0.0667	-0.0720	-0.0561	-0.0628
	(0.128)	(0.142)	(0.129)	(0.142)
edu3	-0.0678	0.0820	-0.0745	0.0859
	(0.0582)	(0.0644)	(0.0582)	(0.0646)
edu4	-0.0883	0.0951	-0.0965	0.100
	(0.0706)	(0.0784)	(0.0709)	(0.0788)
edu5	0.00808	0.0815	0.00164	0.0913
	(0.0763)	(0.0867)	(0.0767)	(0.0876)
wealth	-0.0806	0.354	-0.117	0.368
	(0.164)	(0.239)	(0.176)	(0.238)
iemp04	0.0630	0.544	.	0.544
	(0.0880)	(0.0956)	.	(0.0947)
iemps04	0.115	0.623	.	0.629
	(0.110)	(0.117)	.	(0.116)
Constant	-1.175	-1.321	-1.193	-1.332
	(0.115)	(0.125)	(0.115)	(0.126)
# of Obs.	3368			
Log-likelihood	-1951.51	-1560.15	-3511.84	

Note: Robust standard errors are reported in brackets

TABLE 7. Estimated parameter for the multivariate finite mixture (Model 4)

	Main System		Auxiliary System						Multinomial Membership Probabilites		
	Hosp. 2006	Ins. 2006	Hosp. 2004	Ins. 2004	Hosp. 2002	Ins. 2002	ADL	DIS	$m=2$	$m=3$	$m=4$
$\alpha(1)$	-1.592	-1.978	-2.118	-2.620	-2.114	-2.293	-2.555	1.751	.	.	.
	(0.107)	(0.119)	(0.146)	(0.200)	(0.138)	(0.145)	(0.172)	(0.0840)	.	.	.
$\alpha(2)$	-1.830	1.470	-1.919	2.647	-1.746	1.523	-2.459	2.027	.	.	.
	(0.181)	(0.119)	(0.154)	(0.231)	(0.134)	(0.120)	(0.182)	(0.0955)	.	.	.
$\alpha(3)$	0.167	1.262	0.175	2.209	0.00716	1.709	-0.554	3.977	.	.	.
	(0.167)	(0.146)	(0.122)	(0.234)	(0.115)	(0.177)	(0.112)	(0.150)	.	.	.
$\alpha(4)$	0.0484	-2.268	0.141	-2.330	-0.147	-1.852	-0.309	4.110	.	.	.
	(0.0954)	(0.183)	(0.101)	(0.213)	(0.0955)	(0.148)	(0.0943)	(0.139)	.	.	.
α_U	.	.	.	.	.	.	.	.	-0.245	-0.392	0.211
	.	.	.	.	.	.	.	.	(0.0838)	(0.137)	(0.121)
spins06	0.328	.	.	.	.	.	.	.	.	.	.
	(0.133)	.	.	.	.	.	.	.	.	.	.
fem	-0.389	-0.188	.	.	.	.	.	.	.	.	.
	(0.225)	(0.271)	.	.	.	.	.	.	.	.	.
age	0.237	-1.434	.	.	.	.	.	.	.	.	.
	(0.678)	(0.843)	.	.	.	.	.	.	.	.	.
age^2	-0.135	1.302	.	.	.	.	.	.	.	.	.
	(0.699)	(0.879)	.	.	.	.	.	.	.	.	.
age^3	0.709	0.501	.	.	.	.	.	.	.	.	.
	(0.865)	(1.062)	.	.	.	.	.	.	.	.	.
fem*age	-0.128	-0.0615	.	.	.	.	.	.	.	.	.
	(0.860)	(1.068)	.	.	.	.	.	.	.	.	.
fem*age^2	0.0496	-0.317	.	.	.	.	.	.	.	.	.
	(0.195)	(0.248)	.	.	.	.	.	.	.	.	.
fem*age^3	-0.0715	-0.0376	.	.	.	.	.	.	.	.	.
	(0.234)	(0.293)	.	.	.	.	.	.	.	.	.
edu3	.	.	.	.	.	.	.	.	0.715	-0.543	-0.289
	.	.	.	.	.	.	.	.	(0.146)	(0.189)	(0.170)
edu4	.	.	.	.	.	.	.	.	0.543	-0.304	-0.487
	.	.	.	.	.	.	.	.	(0.169)	(0.220)	(0.207)
edu5	.	.	.	.	.	.	.	.	-0.0919	-0.0626	-0.338
	.	.	.	.	.	.	.	.	(0.184)	(0.263)	(0.251)
wealth	.	.	.	.	.	.	.	.	1.198	-4.190	-0.620
	.	.	.	.	.	.	.	.	(0.345)	(0.902)	(1.090)

Note: Standard errors are reported in brackets
Note: Cut points coefficients in ADL and DIS are suppressed

Table 8. Estimated parameter for the standard finite mixture model

	Main System		Multinomial Membership Probabilites	
	Hospital 2006	Insurance 2006	$m=2$	$m=2$
spins06	0.483	.	.	.
	(0.483)	.	.	.
spins04	.	.	-2.763	-2.011
	.	.	(0.367)	(0.311)
spins02	.	.	-1.691	-1.562
	.	.	(0.286)	(0.269)
h04	.	.	-0.326	1.455
	.	.	(0.197)	(0.292)
h02	.	.	-0.0480	0.113
	.	.	(0.160)	(0.204)
dis	.	.	-0.00290	0.595
	.	.	(0.0659)	(0.123)
adl	.	.	0.0850	0.314
	.	.	(0.109)	(0.114)
edu3	.	.	-0.218	-0.275
	.	.	(0.175)	(0.239)
edu4	.	.	-0.270	-0.355
	.	.	(0.210)	(0.292)
edu5	.	.	-0.272	-0.162
	.	.	(0.228)	(0.309)
wealth	.	.	-1.690	0.0452
	.	.	(0.613)	(0.599)
fem	-0.248	-0.299	.	.
	(0.282)	(0.508)	.	.
age	0.774	-2.454	.	.
	(0.811)	(1.694)	.	.
age^2	-0.419	2.296	.	.
	(0.805)	(1.710)	.	.
age^3	0.266	1.687	.	.
	(1.029)	(1.980)	.	.
fem*age	0.208	-0.877	.	.
	(0.993)	(1.881)	.	.
fem*age^2	0.109	-0.540	.	.
	(0.219)	(0.465)	.	.
fem*age^3	-0.131	0.111	.	.
	(0.265)	(0.501)	.	.
$\alpha(1)$	-1.218	2.326	.	.
	(0.492)	(0.458)	.	.
$\alpha(2)$	-1.569	-3.227	.	.
	(0.162)	(0.787)	.	.
$\alpha(3)$	11.24	-1.677	.	.
	(89.22)	(0.252)	.	.
α_U	.	.	0.255	-1.414
	.	.	(0.160)	(0.330)

Note: Standard errors are reported in brackets

References

[1] Abbring, J. H., Chiappori, P.-A., Heckman, J. J. and Pinquet, J. (2003). Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish? *Journal of the European Economic Association* 1: 512–521.

[2] Abbring, J. H., Chiappori, P.-A. and Pinquet, J. (2003). Moral hazard and dynamic insurance data. *Journal of the European Economic Association* 1: 767–820.

[3] Arnott, R. J. and Stiglitz, J. E. (1990). The basic analytics of moral hazard. NBER Working Papers 2484, National Bureau of Economic Research, Inc.

[4] Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review* 53: 941–973.

[5] Bartholomew, D. J. and Knott, M. (1999). *Latent variable models and factor analysis: a unified approach*. London:Arnold.

[6] Bergsma, W., Croon, M. and Hagenaars, J. A. (2009). *Marginal models: For dependent, clustered, and longitudinal categorial data*. Springer Verlag.

[7] Buchmueller, T. C., Couffinhal, A., Grignon, M. and Perronnin, M. (2004). Access to physician services: does supplemental insurance matter? evidence from france. *Health Economics* 13: 669–687.

[8] Cameron, A. C. and Trivedi, P. K. (1991). The role of income and health risk in the choice of health insurance: Evidence from australia. *Journal of Public Economics* 45: 1 – 28.

[9] Cameron, A. C., Trivedi, P. K., Milne, F. and Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in australia. *The Review of Economic Studies* 55: 85–106.

[10] Cameron, S. V. and Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *The Journal of Political Economy* 106: 262–333.

[11] Cardon, J. H. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics* 32: 408–427.

[12] Carlin, C. and Town, R. (2010). Adverse selection: The dog that didn't bite. *Manuscript. University of Minnesota* .

[13] 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: 499–511.

[14] Chiappori, P.-A., Jullien, B., Salanié, B. and Salanié, F. (2006). Asymmetric information in insurance: General testable implications. *The RAND Journal of Economics* 37: 783–798.

[15] Chiappori, P.-A. and Salanié, B. (2000). Testing for asymmetric information in insurance markets. *The Journal of Political Economy* 108: 56–78.

[16] Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *Review of Economics and Statistics* 87: 197–207.

[17] Cohen, A. and Spiegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance* 77: 39–84.

[18] Crocker, K. J. and Snow, A. (1986). The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy* 94: pp. 321–344.

[19] Cutler, D. M., Finkelstein, A. and McGarry, K. (2008). Preference heterogeneity and insurance markets: explaining a puzzle of insurance. *American Economic Review* 98:

157–162.

[20] De Meza, D. and Webb, D. C. (2001). Advantageous selection in insurance markets. *The RAND Journal of Economics* 32: 249–262.

[21] Deb, P. and Trivedi, P. K. (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of Applied Econometrics* 12: 313–336.

[22] 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: 601–625.

[23] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 1–38.

[24] Eichner, M. J. (1998). The demand for medical care: What people pay does matter. *The American Economic Review* 88: 117–121.

[25] Einav, L. and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives* 25: 115–38.

[26] Einav, L., Finkelstein, A. and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *The Quarterly Journal of Economics* 125: 877–921(45).

[27] Einav, L., Finkelstein, A. and Levin, J. (2010). Beyond testing: Empirical models of insurance markets. *Annual Review of Economics* 2: 311–336.

[28] Einav, L., Ryan, S., Finkelstein, A., Schrimpf, P. and Cullen, M. (2011). Selection on moral hazard in health insurance. *working paper* .

[29] Ettner, S. L. (1997). Adverse selection and the purchase of Medigap insurance by the elderly. *Journal of Health Economics* 16: 543–562.

[30] Fang, H., Keane, M. P. and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy* 116: 303–350.

[31] Finkelstein, A. and McGarry, K. (2006). Multiple dimensions of private information: Evidence from the long-term care insurance market. *The American Economic Review* 96: 938–958.

[32] Finkelstein, A. and Poterba, J. (2006). Testing for adverse selection with "unused observables". Working Paper 12112, NBER.

[33] Forcina, A. (2008). Identifiability of extended latent class models with individual covariates. *Computational Statistics and Data Analysis* 52: 5263–5268.

[34] Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* 61: 215–231.

[35] Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52: 271–320.

[36] Hemenway, D. (1990). Propitious selection. *The Quarterly Journal of Economics* 105: 1063–1069.

[37] 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: 513 – 522.

[38] Hoy, M. (1982). Categorizing risks in the insurance industry. *The Quarterly Journal of Economics* 97: 321–36.

[39] Huang, G.-H. and Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika* 69: 5–32.

[40] Hurd, M. D. and McGarry, K. (1997). Medical insurance and the use of health care services by the elderly. *Journal of Health Economics* 16: 129–154.

[41] Jones, A. M., Koolman, X. and Van Doorslaer, E. (2007). The impact of supplementary private health insurance on the use of specialists in European countries. In *Annales d'Economie et de Statistiques*, *83*.

[42] Jong, R. M. de and Woutersen, T. (2011). Dynamic time series binary choice. *Econometric Theory* 27: 673–702.

[43] Joreskog, K. G. and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* 70: 631–639.

[44] Khandker, R. K. and McCormack, L. A. (1999). Medicare spending by beneficiaries with various types of supplemental insurance. *Medical Care Research and Review* 56: 137.

[45] Lustig, J. (2011). Measuring welfare losses from adverse selection and imperfect competition in privatized medicare. *Manuscript. Boston University* .

[46] Matzkin, R. (2007). Nonparametric identification. *Handbook of Econometrics* 6: 5307–5368.

[47] McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley-Interscience.

[48] Monfardini, C. and Radice, R. (2008). Testing exogeneity in the bivariate probit model: A monte carlo study. *Oxford Bulletin of Economics and Statistics* 70: 271–282.

[49] Olivella, P. and Vera-Hernandez, M. (2012). Testing for asymmetric information in private health insurance. *The Economic Journal* Forthcoming.

[50] Pauly, M. V. (1968). The economics of moral hazard: Comment. *The American Economic Review* 58: 531–537.

[51] Propper, C. (1993). Constrained choice sets in the u.k. demand for private medical insurance. *Journal of Public Economics* 51: 287 – 307.

[52] Rothschild, C. (2011). The efficiency of categorical discrimination in insurance markets. *Journal of Risk and Insurance* 78: 267–285.

[53] Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics* 90: 629–649.

[54] Schellhorn, M. (2001). The effect of variable health insurance deductibles on the demand for physician visits. *Health Economics* 10: 441–456.

[55] Smart, M. (2000). Competitive insurance markets with two unobservables. *International Economic Review* 41: 153–69.

[56] Villeneuve, B. (2003). Concurrence et antisélection multidimensionnelle en assurance. *Annales d'Economie et de Statistique* : 06.

[57] Wambach, A. (2000). Introducing heterogeneity in the rothschild-stiglitz model. *The Journal of Risk and Insurance* 67: 579–591.

[58] Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69: 309 – 312.

[59] Zweifel, P. and Manning, W. G. (2000). Moral hazard and consumer incentives in health care. In Culyer, A. J. and Newhouse, J. P. (eds), *Handbook of Health Economics*. Elsevier, Handbook of Health Economics *1*, chap. 8, 409–459.