Health, Econometrics and Data Group

THE UNIVERSITY of York

HEDG Working Paper 08/26

Testing for Asymmetric Information in Insurance Markets with Unobservable Types

Valentino Dardanoni

Paolo Li Donni

October 2008

ISSN 1751-1976

TESTING FOR ASYMMETRIC INFORMATION IN INSURANCE MARKETS WITH UNOBSERVABLE TYPES

VALENTINO DARDANONI AND PAOLO LI DONNI

Abstract. In two important recent papers, Finkelstein and McGarry [25] and Finkelstein and Poterba [28] propose a new test for asymmetric information in insurance markets that considers explicitly unobserved heterogeneity in insurance demand. In this paper we propose an alternative implementation of the Finkelstein-McGarry-Poterba test based on the identification of unobservable types by use of finite mixture models. The actual implementation of our test follows some recent advances on marginal modelling as applied to latent class analysis; formal testing procedures for the null of asymmetric information and for the hypothesis that private information is indeed multidimensional can be performed by imposing restrictions on the behavior of these unobservable types. To show the potential applicability of our approach, we look at the long term insurance market as analyzed in Finkelstein and McGarry [25], where we also find strong evidence for both asymmetric information and multidimensional unobserved heterogeneity.

JEL Classification Numbers D82, G22, I11
Keywords Asymmetric Information, Unobservable Types, Latent Class Analysis, Long Term Insurance Market.

1. Introduction

The effects of asymmetric information for the efficient operation of insurance markets has been one of the most actively research topics in economics in the past thirty years, starting from the classic Rothschild and Stiglitz [38] paper. On the other hand, empirical research on the presence of asymmetric information in specific insurance markets is rather less established and the evidence whether asymmetric information exists in specific insurance markets is not yet really settled.[1] Much empirical research has used the so called "positive correlation test" of Chiappori and Selanié [13], which rejects the null of absence of asymmetric information in a given insurance market when, conditional on insurer's characteristics used by companies to price contracts,

Date: October 23, 2008.

[1]See e.g Barsky *et al.* [1], Ettner [22], Hurd *et al.* [32], Philipson and Becker [36], Mitchell *et al.* [35], Cawley and Philipson [10], Chiappori and Salanié [13], Chiappori [11], Cardon and Hendel [8], de Meza and Webb [19], Dionne *et al.* [21], Finkelstein and Poterba [26], Bernheim *et al.* [6], McCarthy and Mitchell [34], Brown and Finkelstein [7], Finkelstein and Poterba [27], Cohen [14], Cohen and Einav [15], Chiappori *et al.* [12], Fang *et al.* [23], for a small sample of the growing empirical literature on asymmetric information in various insurance markets.

individuals with more insurance experience more of the insured risk. In two important recent papers, Finkelstein and McGarry [25] and Finkelstein and Poterba [28] (respectively FM and FP henceforth), after observing that the positive correlation test runs into difficulties when individuals have private information not only about their risk class but also about their risk preferences and other characteristics which affect insurance demand, propose and implement a new test for asymmetric information that considers explicitly unobserved heterogeneity in insurance demand. In particular, the null hypothesis of absence of asymmetric information can be rejected if, conditionally on insurer's characteristics used to price insurance contracts, there are other observable characteristics that are correlated with both insurance coverage and ex post risk occurrence. Typical examples of observable variables which are not used to price insurance contracts but may potentially affect both insurance choice and risk occurrence, include, for example, wealth, occupation, and risk reducing or increasing behavior such as preventive care and smoking. The wide applicability of the Finkelstein-McGarry-Poterba test (henceforth called the FMP test) is witnessed in Cutler, Finkelstein and McGarry [16] (CFM henceforth), where asymmetric information is tested in five different insurance markets (namely life, acute private health, annuities, long-term care, and Medigap), using smoking and drinking behavior, mortality occupational risk, use of preventive care and seat belts as observables which are not used to price insurance contracts but are good proxies for unobservable risk attitudes and true risk class.

In this paper we propose an alternative implementation of FMP test based on the identification of unobservable types by use of finite mixture models. In particular, we consider using the set of observable variables which would be typically used in FMP test as *indicators* for the unobservable risk attitude and true risk class which affect both insurance choice and ex post risk occurrence. The actual implementation of our test follows some recent advances of marginal modelling (Bergsma and Rudas [5] and Bartolucci, Colombi and Forcina [2]) as applied to latent class analysis (Huang and Bandeen-Roche [31], Bartolucci and Forcina [4] and Dardanoni, Forcina and Modica [18]). These methods allow us to identify a finite number of unobservable "types" which have heterogeneous risk attitudes and belong to different true risk classes; formal testing procedures for the null of asymmetric information and for the hypothesis that private information is indeed multidimensional (as forcibly argued by FM, FP and CFM) can then be performed by imposing restrictions on the behavior of these unobservable types.

To show the potential applicability of our approach, we look at the long term insurance market as analyzed in FM. Confirming FM's results, we find that there is strong evidence for both asymmetric information and multidimensional unobserved heterogeneity.

2. FMP Test for asymmetric information

FM and FP consider the following model:

$$I = I(\boldsymbol{x}, r_a, r_c, \epsilon_i)$$

$$O = O(\boldsymbol{x}, r_a, r_c, \epsilon_o)$$

where I denotes purchase of insurance and O denotes actual risk occurrence, which are assumed functions of a vector of observable characteristics $\boldsymbol{X}$ which are used by insurance companies to place the buyer into a risk class, R_a and R_c denote respectively unobservable individual's risk attitude and true risk class, and ϵ_i, ϵ_o denote uncorrelated errors. The basic premise of the FMP test is that there is no asymmetric information if R_a, and R_c are *ignorable* in the two conditional probabilities below (we assume here that I and O are actually binary variables, but the FMP test holds with appropriate modifications in more general cases):

$$\begin{aligned} Pr(I = 1 \mid \boldsymbol{x}, r_a, r_c) &= Pr(I = 1 \mid \boldsymbol{x}) \\ Pr(O = 1 \mid \boldsymbol{x}, r_a, r_c) &= Pr(O = 1 \mid \boldsymbol{x}) \end{aligned} \tag{1}$$

Since R_a, and R_t are not directly observed, the FMP test is actually implemented by searching for any candidate unused observed variable which proxies for either R_a or R_c, which we denote by Z. The FMP test is then based on the following null hypothesis of ignorability of Z:

$$\begin{aligned} Pr(I = 1 \mid \boldsymbol{x}, z) &= Pr(I = 1 \mid \boldsymbol{x}) \\ Pr(O = 1 \mid \boldsymbol{x}, z) &= Pr(O = 1 \mid \boldsymbol{x}). \end{aligned} \tag{2}$$

3. Testing for asymmetric information with unobservable types

An alternative implementation of the FMP test (1) can be obtained by using a finite mixture model which identifies the unobservable "types" which represent different combinations of risk attitude and true risk class. Since finite mixture models allow both the marginal distribution of the types and the distribution of the responses conditional on the types to be unconstrained (with the only limitation that the number of types is finite), assuming that (R_a, R_c) are discrete random variables, we can always rewrite (R_a, R_c) as a random variable U taking values in, say, $\{1, \dots, m\}$, which define m unobservable heterogeneous "types".[2] Thus, when no additional structure is imposed on the distribution of the types, they can capture the multidimensional nature of private information which is has been convincingly stressed by FM, FP and CFM. However, as explained for example by Bartolucci and Forcina [3], the *monotonicity* assumption implies that, after a suitable reordering of the types, the conditional expectation of each response variable is an increasing function of the types' distribution; hence, if the monotonicity assumption holds, one cannot exclude the presence of an unique underlying unobserved variable representing types in some increasing order.

Given the distribution of unobservable types U, equation (1) can be rewritten as

$$\begin{aligned} Pr(I = 1 \mid \boldsymbol{x}, u) &= Pr(I = 1 \mid \boldsymbol{x}) \\ Pr(O = 1 \mid \boldsymbol{x}, u) &= Pr(O = 1 \mid \boldsymbol{x}). \end{aligned} \tag{3}$$

Now, suppose we have a set of observable *indicators* of risk attitude/class such as any of the candidate unused observed variable Z, and suppose also we have k of these indicators $Z_1, \dots, Z_k$. We assume that they are actually binary variables (this restriction aims to simplify the discussion, but our analysis can be performed as long as these indicators are discrete). Let $\boldsymbol{Z}$ denote the vector of these k indicators. Classical latent class analysis (see e.g. the seminal paper by Goodman [29]) tries to identify U by using the vector of binary responses $\boldsymbol{Y} = [I, O, \boldsymbol{Z}]$, by exploiting the

[2] For example, if R_a takes, say, three possible values describing three different attitudes towards risk, and R_c takes four possible values describing four different true risk classes, then there are $m = 12$ possible unobservable "types" described by U. In practice, unobservable types which are not sufficiently distinct in the data are actually clumped together.

so-called *local independence* assumption, which states that the unobservable variable U makes observed responses conditionally independent. If a variable U which defines a set of unobservable types and makes $\boldsymbol{Y}$ conditionally independent could be identified, then U would capture unobservable risk attitudes and classes since knowing U would imply, say, that knowledge of any observable variable Z_i would be irrelevant for predicting insurance purchase I, ex post risk occurrence O or any other indicator Z_j, and vice versa. In other words, under local independence, responses can be considered as observable manifestations of underlying true characteristics and preferences.

However, classical latent class models are not useful in our specific context, and in particular the assumption of local independence may be too restrictive for this application. The first point to notice is that equation (3) implies that insurance choice and risk occurrence depend not only on U but also on the set of covariates $\boldsymbol{X}$ used by insurance companies to classify individuals in different risk classes; moreover, it seems likely that the probability of belonging to any given unobservable type may also depend on $\boldsymbol{X}$. Secondly, we may want to allow some residual association between I and O even after conditioning on the unobservable heterogeneity U. Thus, we use recent advances in latent class analysis with discrete responses which allow for these extensions. In particular, Huang and Bandeen-Roche [31] explain how a discrete response finite mixture model can be identified and estimated in the presence of covariates, under the local independence assumption, Bartolucci and Forcina [4] extend it by allowing residual association on responses in the context of capture/recapture models, and Dardanoni, Forcina and Modica [18] apply a model similar to that considered in this paper to study the direct effect of parents' schooling on children's scholastic achievements.

Suppose we have data on a set of k indicators for risk attitudes/classes, a vector of covariates $\boldsymbol{x}$ which describes the variables used by insurance companies to price insurance contracts, and the two binary responses I and O; the proposed model can be written as a multivariate logit model with fixed effects as:

$$\begin{aligned}
&\lambda_I = \alpha_I(u) + \boldsymbol{x}'\boldsymbol{\beta}_I \\
&\lambda_O = \alpha_O(u) + \boldsymbol{x}'\boldsymbol{\beta}_O \\
&\lambda_{Z_i} = \alpha_{Z_i}(u),\ i = 1, \dots, k \\
&\lambda_U(u) = \alpha_U(u) + \boldsymbol{x}'\boldsymbol{\beta}_U(u),\ u = 1, \dots, m-1 \\
&\lambda_{IO} = \alpha_{IO}
\end{aligned} \tag{4}$$

where λ_{IO} denotes the *log-odds ratio* between I and O which captures the residual association between insurance purchase and risk occurrence, and $\lambda_U(u) = \log\left(\frac{Pr(U=u+1)}{Pr(U=u)}\right)$ denote the set of consecutive logits.

Within model (4),

- the null hypothesis of absence of asymmetric information (that is, equation 2) can be tested by imposing the restriction that the fixed effects $a_I(u)$ and $a_{NH}(u)$ are constant across types U. This can be implemented with a standard LR test which has asymptotic chi-squared distribution with $2 \times (m-1)$ degrees of freedom in the presence of m types;
- the null hypothesis that there is a underlying *unidimensional* unobservable variable U such that choices are monotonically dependent on it can be tested by setting a system of linear inequalities as explained for example in Bartolucci and Forcina [3]. 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 [30] for a general exposition, Dardanoni and Forcina [17] for an explanation of how the mixing weights can be calculated by simulations, and Kodde and Palm [33] for bounds on the test distribution).

4. An application to long-term care insurance

FM study the long-term insurance market in the USA, and in particular the relationship between insurance purchase and subsequent nursing home use. Long-term care expenditure risk is one the greatest financial risks faced by the elderly in US; to get a quantitative feeling of its importance, the amount of expenditure in nursing home care in 2004 was about 1.2% of the US GDP. Furthermore, as argued by FM, it is a good market to study since it is not heavily regulated. Their data comes from the Health and Retirement Survey (HRS); the average age of respondent is 78, and about 11% of individuals in the sample have long-term care insurance (a binary variable which we denote by I) in 1995; 16% of the individuals in the sample enter a nursing home (a binary variable which we denote by NH) in the following 5 years period.

FM notice that in the sample there is *negative* correlation between insurance purchase and nursing home use. They also perform the Chiappori and Salanié test for asymmetric information, conditioning on risk classification by calculating, by means of a standard actuarial model,[3] the probability of nursing home use as estimated by insurance companies. FM perform the Chiappori and Salanié test by using both a bivariate probit and two single equation probits, and they always find conditional *negative* association. The same results are found using the subsample which excludes the chronically ill and includes only the wealthier individuals.

As an explanation of these puzzling results, FM argue that there are two conflicting sources of private information, namely individual's true risk class and risk attitudes; individuals who are higher risk tend to both buy more insurance and use more the nursing home, while individuals who are more risk averse tend to buy more insurance but are less likely to use the nursing home. Therefore, two types of people buy insurance: individuals with private information that they are high risk and individuals with private information that they have high risk aversion. Ex post, the former are higher risk then predicted; the latter are lower risk. In aggregate, ex post those who buy more insurance are not higher nursing home users. In terms of the FMP test (2), the absence of asymmetric information is formally rejected since various unused observable proxies Z such as preventive activities, wealth or seat belt use affect both insurance purchase and nursing home use.

[3]They alternatively use as controls a rich set of covariates typically used by insurance companies to price contracts, but their results do not change significantly.

5. Asymmetric information in long-term care insurance with unobservable types

We implement the test (3) by using FM dataset from the Health and Retirement Survey as reported in table 3 of their paper (FM [25] pg. 946).[4] As indicators for U we use six binary variables: Drinking (DR) which takes the value 1 if the subject has less than three drinks per day; Smoking (SM) which takes value 1 if the subject currently does not smoke; Seat Belt (SB) which takes value 1 if the subject always wears seat belts; Preventive Care (PC) which takes value 1 if the subject has taken any gender appropriate preventive care procedures in the past year; Subjective Probability (SP) which takes value 1 if the subject believes with positive probability that she will use a nursing home in the following 5 years; Wealth (WE) which takes value 1 if the subject belongs to the highest wealth quartile. In the sample 96% of respondents report no drinking problem, 77% always use seat belt, 91% do not smoke, 28% are in the top wealth quartile, 51% believe with positive probability that they will enter a NH in the future, and 94% undertook some gender-appropriate preventive health care procedure. As covariates $\boldsymbol{X}$ we use the insurance company's estimate of the probability of using the nursing home, which is calculated by FM from a standard actuarial model. We create 10 risk categories by considering deciles, so $\boldsymbol{X}$ is actually a vector of 9 dummies.

Parameters in model (4) are estimated by the EM algorithm; a technical appendix discusses estimation and identification of the model.[5]

6. Results

The first issue when dealing with finite mixture models is to determine the number of types U. The standard approach in the literature is to use an information criterion such as Akaike Information Criteria (AIC) or Schwartz's Bayesian Information Criteria (BIC) which penalize the likelihood for parameters' proliferation. Calculating the AIC and BIC values for model (4) (reported in table 1 below), in this sample we arrive at the conclusion that three types are adequate to model the unobservable

[4]FM's dataset is available in the AER website. We thank FM, and the AEA for their policy of providing data for published articles.

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

heterogeneity. With three types, simulations (as explained in the appendix) reveal that parameters are solidly identified.

Table 1. Model

	2LC	3LC	4LC
LL	-15055	-14989	-14987
BIC	30481	30503	30650
AIC	30197	30102	30133

We then test the presence of residual association of I and NH; the log-odds ratio is equal to 0.053 with a standard error of 0.173. Thus we conclude that in this sample there is no evidence of residual association between insurance purchase and nursing home use, as confirmed by the LR test which has a test statistic equal to 0.079 with 1 df (p-value 0.778).

The final model we estimate has 62 parameters: 36 regression coefficients $\boldsymbol{\beta}_I$, $\boldsymbol{\beta}_{NH}$, $\boldsymbol{\beta}_U(U = 1)$ and $\boldsymbol{\beta}_U(U = 2)$; 6 fixed effects $a_I(u)$ and $a_{NH}(u)$; 18 logits of the conditional probabilities for the indicators; and 2 intercepts for the marginal probability for U. Parameters estimates and their standard errors are reportedo in Appendix B. Notice that in this simpler model without residual association we can apply Theorem 1 in Huang and Bandeen-Roche [31] to establish parameters' identification without resorting to simulations.

Table 2 below reports the conditional probabilities for the six indicators for the three types, and Table 3 below shows the estimated probabilities of purchasing insurance and using a nursing home for the ten risk classes and the three types. A glance at the tables reveals substantial heterogenous attitudes and behaviors towards risky activities by the three types. However, there seems to be a natural ordering of the three types in terms of their cautiousness, such that, going from "types 1" to "types 2" and then to "types 3", there is a significant increase in the probability of using seat belts and preventive care, of refraining from smoking, and believing that one may need a nursing home in the near future with positive probability. The probability of eventually not using the nursing home is also increasing in U.

On the other hand, this monotonic pattern does not hold in this sample for the drinking behavior and wealth indicators, and also for the probability of buying insurance. The question naturally arises then whether these apparent violations of monotonicity are due to sampling variations, or rather to the presence of more than one underlying unobservable variable which pull U in different directions. As discussed above, techniques of order restricted inference can be employed to test the monotonicity assumption; since U has three levels, the assumption that the conditional probability of a given response is monotonically increasing in U involves imposing two inequality constraints. The LR test statistic for the monotonicity null is equal to 1.973, 2.301 and 6.343 respectively for the DR, WE and I. The conservative 5% critical value (Kodde and Palm ([33], page 1246) is equal to 5.138; thus, while violations of monotonicity in the two indicators DR and WE are small enough that they can be assumed to be due to sampling variation, the non monotonic pattern of insurance purchase suggests the presence of more than one underlying unobservable variables which have contrasting effects on the decision to purchase insurance. In particular, our results are in accord with the analysis of FM who found insurance purchase and risk occurrence depend on private information on *both* cautiousness and risk type, which then operate in offsetting directions.

A glance at table 5 reveals also that the odds of being a type 2 rather than 1 increase with the risk class used by insurance company (the vector $\boldsymbol{\beta}_U(U = 1)$ is increasing in $\boldsymbol{X}$), while the odds of being a type 3 rather than 2 individual decrease with it (the vector $\boldsymbol{\beta}_U(U = 2)$ is decreasing). Thus the probability of being a given type U seem to also follow a non monotonic pattern with respect to the risk classification by the insurance company, whereas the probability of being an unobservable type $U = 1, 2, 3$ first increases and then decreases with being classified as a high risk by the insurance company.

Table 2. Indicators' conditional probabilities

	DR	SM	SB	PC	SP	WE
$U = 1$	0.837	0.627	0.633	0.825	0.361	0.189
$U = 2$	0.993	0.937	0.664	0.912	0.470	0.118
$U = 3$	0.978	0.968	0.893	0.994	0.564	0.464

We then turn our attention to testing the null of absence of asymmetric information (equation 3). The LR test statistic is equal to 62.83, which is asymptotically distributed as a chi-square with 4 degrees of freedom (p-value 0.0001); therefore the null of absence of asymmetric information is soundly rejected in this sample.

In particular, returning on Table 3 below, it seems that the three heterogeneous types behave quite differently regarding to their probabilities of purchasing insurance (type 3 individuals on average have almost ten times as much probability of buying insurance than type 1's and almost twice as much probability than type 2's) and their probability of future nursing home use (type 1 individuals on average have 1.7 times as much probability of using the nursing home than type 3's, and 1.3 times as much probability than type 2's).

TABLE 3. Insurance and nursing home conditional probabilities

	$I1$	$I2$	$I3$	$I4$	$I5$	$I6$	$I7$	$I8$	$I9$	$I10$	avg.
$U=1$	0.099	0.108	0.090	0.078	0.114	0.096	0.096	0.105	0.076	0.137	0.096
$U=2$	0.025	0.027	0.022	0.019	0.029	0.024	0.024	0.026	0.018	0.035	0.024
$U=3$	0.191	0.206	0.174	0.152	0.215	0.186	0.184	0.199	0.149	0.253	0.184

	$NH1$	$NH2$	$NH3$	$NH4$	$NH5$	$NH6$	$NH7$	$NH8$	$NH9$	$NH10$	avg.
$U=1$	0.056	0.080	0.097	0.120	0.177	0.184	0.261	0.383	0.356	0.498	0.190
$U=2$	0.041	0.059	0.072	0.090	0.134	0.140	0.204	0.310	0.286	0.418	0.148
$U=3$	0.030	0.043	0.053	0.066	0.101	0.105	0.156	0.245	0.224	0.342	0.114

Thus, our results confirm the apparently puzzling "favorable selection" phenomenon which is reported in many other empirical insurance markets investigations, which concerns the existence of individuals who are both more likely to buy insurance and less likely to incur risk they have bought insurance for; as eloquently argued by FM, FP and CFM, the root of this apparent phenomenon may be precisely the multidimensionality of private information detected above.

Finally, notice that while nursing home use is strongly correlated with the risk class in which insurance companies put individuals (see the vector $\boldsymbol{\beta}_{NH}$ in table 5 in Appendix B), there is no apparent relation between insurance purchase and risk class; the LR test statistic for equality of the insurance purchase probabilities across classes (constancy of $\boldsymbol{\beta}_I$ in table 5 across deciles) is equal to 11.141 with 9 degrees of freedom (p-value 0.2662). This result seems plausible since higher risk class individuals on the one hand have higher probability of using the service which they have bought insurance for, but on the other hand have to pay a greater insurance premium for it.

7. Appendix A: Estimation and Identification of Model (4)

7.1. Likelihood.

7.1.1. *Incomplete data (observable) likelihood.* Let $\boldsymbol{Y} = [I, NH, DR, SM, SB, PC, SP, WE]$ denote the 8 observable binary response variables employed in model (4) and $\boldsymbol{y}$ any of the 2^8 possible response configuration. Let then $q_{\boldsymbol{y}} = Pr(Y_1 = y_1, \dots Y_8 = y_8)$, denote the corresponding cell probabilities, which may be arranged into the 2^8-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, and let $\boldsymbol{n}$ be the 2^8-sized vector containing the observed frequency table of $\boldsymbol{Y}$, arranged in the same lexicographic order as $\boldsymbol{q}$. The (kernel of) the log-likelihood can then be written as

$$L(\boldsymbol{n}, \boldsymbol{q}) = \boldsymbol{n}' \log(\boldsymbol{q}).$$

7.1.2. *Complete data (unobservable) likelihood.* Recall that $(u, \boldsymbol{y})$ denotes any of the $m \cdot 2^8$ possible (unobservable) response configurations, and $\boldsymbol{p}$ arrays the unobservable joint distribution of $(U, \boldsymbol{Y})$. Let then $\boldsymbol{m}$ be the $(m \cdot 2^8)$-sized vector containing the unobservable frequency table of $(U, \boldsymbol{Y})$, arranged in the same lexicographic order as $\boldsymbol{p}$. The (kernel of) the complete data log-likelihood can then be written as

$$\Lambda(\boldsymbol{m}, \boldsymbol{p}) = \boldsymbol{m}' \log(\boldsymbol{p}).$$

7.2. The maximization of the complete data log-likelihood.

Maximization of the unobservable log-likelihood can be seen as a problem of incomplete data which may be tackled by the EM algorithm (Dempster, Laird and Rubin [20]).

7.2.1. *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 U given the observed response configuration $\boldsymbol{y}$, $Pr(U = u \mid \boldsymbol{y}) = p_{u,\boldsymbol{y}}/q_{\boldsymbol{y}}$, so that $m(u, \boldsymbol{y}) = n(\boldsymbol{y}) Pr(u, \boldsymbol{y})$ by the expectation of the multinomial distribution for U, where $m(u, \boldsymbol{y})$ and $n(\boldsymbol{y})$ denote the frequency of observations with responses respectively given by $(u, \boldsymbol{y})$ and $\boldsymbol{y}$.

7.2.2. *The M step.* Maximization of Λ with respect to the model parameters can be implemented with the method of scoring. As explained for example in Bartolucci and Forcina [4] and Dardanoni, Forcina and Modica [18], following the marginal modelling approach of Bergsma and Rudas [5] and Bartolucci, Colombi and Forcina [2], we can define an invertible and differentiable mapping from the (unobservable) joint distribution of $(U, \boldsymbol{Y})$ to the vector of logits described in model (4) which is linearly linked to the model parameters. In particular,

(1) rewrite model (4) as a multivariate linear logit system

$$\boldsymbol{\lambda} = \boldsymbol{D}(\boldsymbol{x})\boldsymbol{\psi}$$

where where $\boldsymbol{D}(\boldsymbol{x})$ is a design matrix whose dependence on $\boldsymbol{x}$ reflects the effect of the covariates on the different elements of the joint distribution, and $\boldsymbol{\psi}$ is the vector which collects the model parameters α's and β's;

(2) the mapping from $\boldsymbol{p}$ to $\boldsymbol{\lambda}$ can be written in explicit form by constructing an appropriate contrast matrix $\boldsymbol{C}$ (whose rows have elements summing to zero) and a marginalization matrix $\boldsymbol{M}$ (a matrix made of 1's and 0's) such that

$$\boldsymbol{\lambda} = \boldsymbol{C}\ln(\boldsymbol{M}\boldsymbol{p}),$$

and Theorem 1 in Bartolucci–Colombi–Forcina [2] shows that this mapping is invertible and differentiable for any $\boldsymbol{p}$ with strictly positive elements.

Given the mappings from $\boldsymbol{p}$ to $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}$ to $\boldsymbol{\psi}$, the score vector and the expected information matrix required for the computation are then obtained by the chain rule.

7.3. **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. However, the correct information matrix may be computed from the complete data log-likelihood as explained for example in Bartolucci and Forcina [4] and Dardanoni, Forcina and Modica [18].

7.4. **Identifiability.** Unfortunately there is no general result in the literature on finite mixture models which can be applied to show that our general model is identifiable. Results of Rothenberg [37] and Catchpole and Morgan [9] indicate that the model is identified whenever the Jacobian of the mapping between the observable joint distribution $\boldsymbol{q}$ and the model parameters $\boldsymbol{\psi}$ is of full rank. Following the methods suggested by Forcina [24], identifiability of model (4) can then be checked by sampling a reasonable number of canonical parameters of the distribution of $\boldsymbol{q}$ and checking the full rank condition of the Jacobian. Since with 10000 runs we could not find a single instance where the rank was any close to be being deficient, we have good practical evidence that our model is indeed identifiable for a wide range of the parameters' space. However, as argued above, if residual association between I and NH is zero, then our model becomes a special case of Huang and Bandeen-Roche [31], and identifiability follows from their Theorem 1.

8. Appendix B: Parameters' estimates

Table 4. Estimated intercepts of model 4

	Coef.	S.E.
Latent variable U		
$\alpha_U(U=1)$	2.7257	0.8471
$\alpha_U(U=2)$	-1.7163	0.4582
Drinking problem		
$\alpha_{DR}(U=1)$	1.6426	0.2438
$\alpha_{DR}(U=2)$	4.9505	1.0105
$\alpha_{DR}(U=3)$	3.7982	0.2688
Seat belt use		
$\alpha_{SB}(U=1)$	0.5449	0.1625
$\alpha_{SB}(U=2)$	0.6814	0.0880
$\alpha_{SB}(U=3)$	2.1290	0.1554
Smoking		
$\alpha_{SM}(U=1)$	0.5226	0.2660
$\alpha_{SM}(U=2)$	2.7101	0.2517
$\alpha_{SM}(U=3)$	3.4269	0.3273
Top wealth quartile		
$\alpha_{WE}(U=1)$	-1.4527	0.2216
$\alpha_{WE}(U=2)$	-2.0110	0.2124
$\alpha_{WE}(U=3)$	-0.1426	0.0923
Subjective probability to use NH		
$\alpha_{SP}(U=1)$	-0.5674	0.1648
$\alpha_{SP}(U=2)$	-0.1200	0.0716
$\alpha_{SP}(U=3)$	0.2610	0.0640
Preventive care behavior		
$\alpha_{PC}(U=1)$	1.5522	0.2028
$\alpha_{PC}(U=2)$	2.3487	0.1312
$\alpha_{PC}(U=3)$	5.2019	1.0637
Long term insurance		
$\alpha_I(U=1)$	-1.8347	0.4719
$\alpha_I(U=2)$	-3.2967	0.4255
$\alpha_I(U=3)$	-1.0792	0.3813
Nursing home admission		
$\alpha_{NH}(U=1)$	-0.0056	0.2877
$\alpha_{NH}(U=2)$	-0.3281	0.1034
$\alpha_{NH}(U=3)$	-0.6534	0.1959

Table 5. Parameters estimates of risk categories deciles

	$\beta_U(U=1)$		$\beta_U(U=2)$		β_I		β_{NH}	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
1	-3.3391	0.8499	3.2678	0.6587	-0.3632	0.3993	-2.8110	0.2925
2	-2.8484	0.7992	2.9530	0.5499	-0.2677	0.3948	-2.4320	0.2628
3	-2.7158	0.7986	2.8548	0.5384	-0.4749	0.3994	-2.2179	0.2527
4	-2.0044	0.7718	2.4425	0.4683	-0.6330	0.3959	-1.9835	0.2285
5	-1.8368	0.7268	1.9208	0.4384	-0.2161	0.3737	-1.5306	0.1889
6	-1.2754	0.7579	1.9925	0.4344	-0.3971	0.3789	-1.4799	0.1906
7	-0.5548	0.7709	1.4735	0.4175	-0.4051	0.3642	-1.0314	0.1632
8	-1.0140	0.7208	1.4164	0.4197	-0.3082	0.3608	-0.4697	0.1510
9	-0.8846	0.6995	0.9252	0.4169	-0.6583	0.3649	-0.5869	0.1460

Rows report the risk categories deciles. Top decile is omitted.

References

[1] Barsky, R., Juster, T., Kimball, M. and Shapiro, M. (1997): "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study." *Quarterly Journal of Economics*, 112(2), pp. 537-579.

[2] 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**: 691-711.

[3] Bartolucci, F. and Forcina, A. (2005): "Likelihood inference on the underlying structure of IRT models", *Psychometrika*, **30**, pp. 140-159

[4] 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

[5] Bergsma, W. and T. Rudas (2002): "Marginal models for categorical data." *The Annals of Statistics*, 30, pp. 140-159.

[6] Bernheim, D., Forni, L., Gokhale, J. and Kotlikoff, L. (2003): "The Mismatch Between Life Insurance and Financial Vulnerabilities: Evidence from the Health and Retirement Survey," *American Economic Review*, 93(1), pp. 354-365.

[7] Brown, J. and Finkelstein, A. (2004): "Supply or Demand: Why is the Market for Long-Term Care Insurance So Small," *NBER Working Paper* 10782.

[8] Cardon, J. and Hendel, I. (2001): "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics*, 32, pp. 408-427.

[9] Catchpole, E. A., and Morgan, B. J. T. (1997): "Detecting parameter redundancy," *Biometrika*, 84, pp. 187-196.

[10] Cawley, J. and Philipson, T. (1999): "An empirical examination of information barriers to trade in insurance," *American Economic Review*, 89(4), pp. 827-846.

[11] Chiappori, P. A. (2000): "Econometric Models of Insurance under Asymmetric Information." In Georges Dionne, ed. *Handbook of Insurance Economics.* London: Kluwer, pp. 365-394.

[12] Chiappori, P. A., Jullien, B., Salanié, B. and Salanié, F. (2006): "Asymmetric Information in Insurance: General Testable Implications," *Rand Journal of Economics*, 37(4), pp. 783-798.

[13] Chiappori, P. A. and Salanié, B. (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108(1), pp. 56-78.

[14] Cohen, A. (2005): "Asymmetric Information and Learning in the Automobile Insurance Market," *Review of Economics and Statistics*, 87, pp. 197-207.

[15] Cohen, A. and Einav, L. (2005): "Estimating Risk Preferences from Deductible Choice," *Working Paper*, Stanford University.

[16] Cutler, D. M., Finkelstein, A. and McGarry, K. (2008): "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance." *NBER Working Paper*, 13746.

[17] 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.

[18] Dardanoni, V., Forcina, A., and Modica, S. (2008): "Direct causal effects in education transmission", *submitted*.

[19] de Meza, D. and Webb, D. C. (2001): "Advantageous selection in insurance markets," *Rand Journal of Economics*, 32(2), pp. 249-262.

[20] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977): "Maximum likelihood for incomplete data via the EM algorithm", *Journal of the Royal Statistical Society* Series B 39, pp. 1-22.

[21] Dionne, G., Gourieroux, C. and Vanasse, C. (2001): "Testing for evidence of adverse selection in the automobile insurance market: a comment", *Journal of Political Economy*, 109(2), pp. 444-451.

[22] Ettner, S, (1997): "Adverse Selection and the Purchase of Medigap Insurance by the Elderly," *Journal of Health Economics*, 16(5), pp. 499-624.

[23] Fang, H., Keane, M. and Silverman, D. (2006): "Sources of Advantageous Selection:Evidence from the Meidgap Insurance Market," *NBER working paper*, 12289.

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

[25] 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.

[26] Finkelstein, A. and Poterba, J. (2002): "Selection Effects in the Market for Individual Annuities: New Evidence from the United Kingdom," *Economic Journal*, 112(476), pp. 28-50.

[27] Finkelstein, A. and Poterba, J. (2004): "Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market," *Journal of Political Economy*, 112(1), pp. 193-208.

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

[29] Goodman, L. (1974): "Exploratory latent structure analysis using both identifiable and unidentifiable models", *Biometrika*, 61, pp. 215-231.

[30] Gourieroux, C. and Monfort, A. (1995): "Statistics and Econometrics Models", Cambridge: *Cambridge University Press.*

[31] 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.

[32] 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.

[33] Kodde, D. A. and Palm, F. C. (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions", *Econometrica*, 54(5), pp. 1243-1248.

[34] McCarthy, D. and Mitchell, O. (2003): "International Adverse Selection in Life Insurance and Annuities," *NBER Working Paper*, 9975.

[35] Mitchell, O. S., Poterba, J. M., Warshawsky, M. and Brown, J. R. (1999): "New Evidence on the Money's Worth of Individual Annuities," *The American Economic Review*, 89, pp. 1299-1318.

[36] Philipson, T. and Becker, G. (1998): "Old-Age Longevity and Mortality-Contingent Claims," *Journal of Political Economy*, 106, pp. 551-573.

[37] Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, pp. 577-591.

[38] Rothschild, M. and Stiglitz, J. E. (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, pp. 630-649.

Facoltà di Economia, Università di Palermo

E-mail address: vdardano@unipa.it

Facoltà di Economia, Università di Palermo

E-mail address: plidonni@gmail.com