A simple test for private information in insurance markets with heterogeneous insurance demand

Li Gan\textsuperscript{a,b}, Feng Huang\textsuperscript{c,d,*}, Adalbert Mayer\textsuperscript{e}

\textsuperscript{a} Department of Economics, Texas A&M University, United States
\textsuperscript{b} NBER, United States
\textsuperscript{c} Institute for Advanced Research, Shanghai University of Finance and Economics, China
\textsuperscript{d} Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, China
\textsuperscript{e} Department of Economics, Washington College, United States

HIGHLIGHTS

- Propose a simple test for asymmetric information in insurance markets with heterogeneous insurance demand.
- First results using a finite mixture model to disentangle the type of selection, adverse selection or advantageous selection.
- Identify the existence of private information, without using direct evidence of private information.

ARTICLE INFO

Article history:
Received 9 April 2015
Received in revised form 23 September 2015
Accepted 27 September 2015
Available online 9 October 2015

JEL classification:
D82
G22
I11

Keywords:
Individual heterogeneity
Private information
Insurance markets

ABSTRACT

A positive correlation between insurance coverage and ex post risk indicates private information in insurance markets. However, this test fails if agents have heterogeneous risk attitudes. We propose a finite mixture model that conditions on unobserved types who differ in their risks preferences and detects asymmetric information even if heterogeneous risk attitudes exist. Our method identifies the existence of private information, without using direct evidence of private information.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Economic theory suggests that the presence of private – or asymmetric – information has important implications for insurance markets. Adverse selection and moral hazard can lead to a sub-optimal provision of insurance and a decrease in welfare. One indicator for the presence of asymmetric information is a positive correlation between an individual’s risk and the decisions to purchase insurance (Chiappori and Salanié, 2000). The empirical results, however, are mixed. Some studies find evidence of adverse selection (Finkelstein and Poterba, 2002). While some other studies report weak or no evidence of adverse selection (Chiappori and Salanié, 2000). Cohen and Siegelman (2010) give a comprehensive review of related empirical work.

One explanation for failure to detect private information is the presence of heterogeneous preferences for insurance. There may be advantageous selection, which means that more cautious people are not only more inclined to purchase insurance but also more likely to put effort in preventing risk exposures. Finkelstein and McGarry (2006), short F&MG, illustrate this for the market of long-term care insurance. They fail to find evidence for a positive correlation between the risk of entering a nursing home and the
decision of purchasing long-term care insurance. However, at the same time they provide direct evidence for the existence of private information about the individual risk of entering a nursing home. They explain that the presence of asymmetric information is masked by heterogeneous risk attitudes and show that: more cautious and wealthier individuals are more likely to purchase long-term care insurance and less likely to enter a nursing home. An advantage of this method is that an incomplete set of variables that explain the individual heterogeneity is normally sufficient to produce consistent estimates in the insurance demand and risk exposure equations, and to detect private information if it exists. We apply this model to the sample of F&MG. We find that – as predicted – the two types of agents behave differently. Conditional on public information and the type of an individual we obtain a statistically significantly positive correlation between \( \text{ex post} \) risk and the insurance purchases. This provides the evidence of the existence of private information. We confirm the finding of F&MG without relying on direct evidence of private information.

The paper is organized as follows, in Section 2, we describe the data and econometric methods, and present the results. Section 3 concludes.

### 2. Data, methods and results

#### 2.1. Data

We illustrate our estimation procedure by applying it to the data assembled by F&MG where direct evidence for private information is available. F&MG apply an actuarial model used by many insurers to calculate a variable that reflects the company prediction of nursing home use which is used to determine premiums. This company prediction captures the available public information, \( X \). Based on a survey question, F&MG construct a measure of private beliefs about the likelihood of moving into a nursing home. We use the private beliefs as a proxy for private information, \( Z \), capturing some but not all of the private information of individuals. The data also contain information about wealth and proxies for risk attitudes. The proxies for risk attitudes are self-reported seat belt usage and whether individuals undertook preventative healthcare measures, such as flu shots or cancer screenings. For more detailed information about sample and variables see F&MG.

Table 1 displays the descriptive statistics. 11% of the individuals have long-term care insurance in 1995 and 16% enter a nursing home at some point from 1995 to 2000.

#### 2.2. Econometric method

F&MG estimate a bivariate probit model of long-term care insurance holdings and nursing home utilization as follows:

\[
\begin{align*}
\text{NH}^* &= X\beta + u, \quad \text{NN} = 1 \text{ if } \text{NH}^* > 0, 0 \text{ otherwise}, \\
\text{LTCI}^* &= X\delta + \varepsilon, \quad \text{LTCI} = 1 \text{ if } \text{LTCI}^* > 0, 0 \text{ otherwise} \quad (1)
\end{align*}
\]

with

\[
\begin{pmatrix}
\mathbf{u} \\
\mathbf{\varepsilon}
\end{pmatrix} \sim N
\begin{pmatrix}
0 \\
0
\end{pmatrix} \cdot
\begin{pmatrix}
1 & \rho \\
\rho & 1
\end{pmatrix}
\]

where \( \text{NH} \) is a binary variable for nursing home utilization between 1995 and 2000. \( \text{LTCI} \) is a binary variable for long-term care insurance holding in 1995. Omitting the private information in (1) leads to a positive correlation (\( \rho > 0 \)) (Chiappori and Salanié, 2000). The error terms follow the standard bivariate normal distribution. However, if individuals have heterogeneous risk preferences, the correlation between the error terms is no longer indicative of the presence of asymmetric information, but reflects a combination of asymmetric information and heterogeneous taste in insurance. F&MG demonstrate that two types of people purchase insurance: individuals with private information that they are high risk (the B old type) and individuals with that have a strong taste for insurance but with lower risk (the T imid type). In aggregate, those with more insurance are not higher risk. Hence, the standard positive correlation test will fail to detect the presence of private information if we mix individuals with heterogeneous risk preferences.\(^2\)

We propose a simple and intuitive test for asymmetric information, based on the finite mixture model. As pointed out by Deb and Trivedi (1997), the finite mixture model provides a natural representation of heterogeneous preference since each latent class can be seen as a “type” of individual. It can also be seen as a discrete approximation of an underlying continuous mixing distribution, which does not need to be specified. Empirically supported by F&MG and Fang et al. (2008), we will start with the simplest case

---

1 Long-term care insurance allows individuals to insure themselves against the cost associated with entering a long term care facility, such as a nursing home.

2 As shown in Table 3 from F&MG (2006), the standard positive correlation tests are unable to reject the null hypothesis of zero correlation in the long-term care insurance market.
in which individuals are assumed to be divided into two types: the bold and the timid. We model the mixing probabilities (probability of being a certain type) for each individual $i$ as a function of proxy variables for risk attitudes of individuals. We model the probability of being the timid type, $\pi$, as:

$$
\pi_i = \Phi \left( \kappa_0 + \kappa_1 \text{Prev}_i + \kappa_2 \text{Seat}_i + \sum_{j=2}^{4} \kappa_{j+1} W_{j,i} \right). \tag{2}
$$

The variables Prev (whether an individual engages in preventive health measures) and Seat (whether an individual always wears a seatbelt) reflect risk attitudes. The Medicaid program offers a better substitute for private insurance to low wealth individuals. Therefore, wealth (measured by three wealth quartile dummies, $W_j$) is also correlated with the risk preferences of individuals.

The log likelihood function of 2-point finite mixture model is given by:

$$
\text{LnL} = \sum_{i=1}^{n} \ln(\pi_i \Phi_2[(2N_i - 1)X_i \beta^T] + (1 - \pi_i) \Phi_2[(2N_i - 1)X_i \beta^B], (2L_i - 1) X_i \beta^T, (2N_i - 1)(2L_i - 1) \rho^T) \tag{3}
$$

where the mixing probabilities $\pi_i$’s are estimated along with all the other parameters of the model. $\Phi_2(\cdot)$ denotes for the standard bivariate normal cumulative distribution function. To compare the results with F&MG, we maintain their assumptions about the error terms. Ensuring that any differences are due to use of the mixture model. We assume that the error terms in each component are bivariate normally distributed. Resulting in multimodality of the error terms in the finite mixture model.

2.3 Results

We estimate the model described by (3) using maximum likelihood. We do not use private information. The results are presented in Table 2. The top panel illustrates the effects of factors predicting the type of an individual. Overall, 28% of individuals belong to the timid type. Individuals of the timid type are characterized by a higher incidence of preventative activities and seat belt use; they also tend to be wealthier.

The bottom panel displays coefficient estimates for the likelihood of entering a nursing home (NH) and the likelihood of purchasing long-term care insurance (LTCI). We restrict the coefficient for public information and the correlation $\rho$ to be identical for the two types, but allow the constant terms to differ. Therefore, the effects of heterogeneous risk preferences are absorbed into the constant terms. For the LTCI model, the estimated constant for the timid type is equal to $-0.269$ and (statistically significant) larger than the estimated constant for the bold type at $-2.312$. For the NH model, the estimated constant for the timid type is $-2.288$ and significantly smaller than the estimated constant for the bold type at $-1.269$. In other words, our empirical model distinguishes two types of people who exhibit clear differences in their behaviors in the long-term care market. The timid type is more likely to purchase long-term care insurance (predicted likelihood of 41%) than the bold type (predicted for less than 1%) and less likely to enter a nursing home (predicted at 3% vs 19%).

Most importantly, by separating individuals into two types based on their preferences, we are able to obtain clear evidence of private information from the standard test. The estimated correlation between the error terms is positive and statistically significant at 0.621 (s.e. = 0.271). It is important to note that this is achieved even without using any data on private information.

As a further check, we also estimate two specifications utilizing one dimension of private information, the individual prediction of entering nursing homes ($Z$). In the first specification we include $Z$ in both the LTCI equation and the NH equation, but not in Eq. (2). The coefficient for $Z$ is positive in both LTCI and NH equations, showing the importance of private information in determining the decisions to buy insurance and to enter nursing homes. Relative to the model without private information the other coefficient estimates do not change much. The point estimate of the correlation between the two error terms, $\rho$, is slightly reduced to 0.566 (s.e. 0.209). Suggesting that $Z$ may only capture a small portion of the private information. As more of the private information is added to the model the correlation $\rho$ should go to zero. In the second specification, we include $Z$ in Eq. (2) but not in the LTCI and NH equations. As expect, the coefficient for $Z$ in Eq. (2) is positive. The likelihood of being the “timid” type and expecting to enter a nursing home are positively associated. Again, the other coefficient estimates are qualitatively unchanged and we find no evidence contradicting the results of our model. The estimate for $\rho$ is now 0.508, albeit estimated less precisely (s.e. 0.335).

An important specification consideration is the number of types. Following the literature, we assume that individuals are heterogeneous in two types, the bold type and the timid type. Theoretically, it is possible to categorize them into three or more types. The number of classes in a finite mixture model is commonly chosen using an information criteria, such as the AIC and BIC. Note

### Table 2

The 2-point finite mixture model.

<table>
<thead>
<tr>
<th>Type (timid type)</th>
<th>Preventive health activity</th>
<th>Always wear seat belt</th>
<th>Top quartile of assets</th>
<th>3rd Wealth quartile</th>
<th>2nd Wealth quartile</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>0.308***</td>
<td>0.382**</td>
<td>0.852**</td>
<td>0.593***</td>
<td>0.349**</td>
<td>-1.574*** (0.185)</td>
</tr>
<tr>
<td>Individual prediction</td>
<td>1.828*** (0.104)</td>
<td>1.828*** (0.104)</td>
<td>-2.288*** (0.215)</td>
<td>-2.288*** (0.061)</td>
<td>-0.628*** (0.184)</td>
<td>-0.269** (0.188)</td>
</tr>
<tr>
<td>LTCI insurance company prediction</td>
<td>-0.628*** (0.184)</td>
<td>-0.628*** (0.184)</td>
<td>-2.312*** (0.237)</td>
<td>-0.621*** (0.271)</td>
<td>-3558.49</td>
<td></td>
</tr>
<tr>
<td>Individual prediction</td>
<td>-0.269** (0.188)</td>
<td>-2.312*** (0.237)</td>
<td>-0.621*** (0.271)</td>
<td>-3558.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.574*** (0.185)</td>
<td>-2.312*** (0.237)</td>
<td>-0.621*** (0.271)</td>
<td>-3558.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: We limit our sample to individuals without missing information. NH reflects any nursing home use between 1995 and 2000. LTCI reflects long-term care insurance coverage in 1995. Our estimates are weighted using the 1995 household weights. *** Denotes statistical significance at the 1% level. ** Denotes statistical significance at the 5% level.
that the 3-point model requires 9 additional parameters to be estimated. The AIC for the 2-point model is 7348.98, smaller than one for the 3-point model at 7856.75. Not surprisingly, the BIC also favors the 2-point model. Increasing the number of types does not help improve the model.

3. Conclusions

Identifying asymmetric information in the insurance markets is important. This paper proposes and estimates a finite mixture model to identify private information in the presence of heterogeneity consumer types. Several characteristics that are correlated with the unobserved types are used to probabilistically determine to which type a person belongs. Accommodating individual heterogeneity, the model can detect the presence of asymmetric information in the insurance markets even when the direct evidence of private information is not available.

Acknowledgment

Feng Huang acknowledges research support by the National Natural Science Foundation of China (Grant No. 71203133).

References