

# A Simple Test of Private Information in the Insurance Markets with Heterogeneous Insurance Demand

March, 2010

Li Gan, Feng Huang, and Adalbert Mayer<sup>1</sup>

A positive correlation between insurance coverage and ex-post risk can be an indicator for private information in insurance markets. However, this test fails if agents have heterogeneous risk attitudes. We propose a way to condition on unobserved types of individuals who differ in their risks preferences. This makes it possible to detect asymmetric information without direct evidence of private information - even if agents have heterogeneous risk attitudes. We apply our technique to the market for long-term care insurance. Finkelstein and McGarry (2006) provide direct evidence for the existence of private information in this market. At the same time they fail to find a positive correlation between insurance coverage and ex-post risk. Our method indicates the existence of private information, without using direct evidence of private information. Our methodology is applicable to other insurance markets and in markets where proxies for private information are not available.

---

<sup>1</sup> Gan: Department of Economics, Texas A&M University and NBER. [gan@econmail.tamu.edu](mailto:gan@econmail.tamu.edu). Huang: Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, China. [fhuang@swufe.edu.cn](mailto:fhuang@swufe.edu.cn). Mayer: Department of Economics, Texas A&M University. [amayer@econmail.tamu.edu](mailto:amayer@econmail.tamu.edu).

## 1. Introduction

Economic theory suggests that the presence of private – or asymmetric – information has important effects on insurance markets.<sup>2</sup> Adverse selection and moral hazard can lead to a sub-optimal provision of insurance and a decrease in welfare. Therefore, it is important to detect and to quantify the effects of asymmetric information in insurance markets.

One indicator for the presence of asymmetric information is a positive correlation between an individual's risk and the decisions to purchase insurance (after controlling for public information).<sup>3</sup> This indicator does not require to directly observe private information and it has been used to investigate a number of insurance markets. The empirical results, however, are mixed and differ by markets. For example, in a life insurance market, Cawley and Philipson (1999) find that the mortality rate of U.S. males who purchase life insurance is below that of the uninsured, even when controlling for many factors such as income that may be correlated with life expectancy. In an auto insurance market, Chiappori and Salanie' (2000) find that accident rates for young French drivers who choose comprehensive automobile insurance are not statistically different from the rates of those opting for the legal minimum coverage, after controlling for observable characteristics known to automobile insurers.

In contrast, Cohen (2005), using data from Israel, finds that new auto insurance customers choosing a low deductible tend to have more accidents, leading to higher total losses for the insurer. In an annuity insurance market, Finkelstein and Poterba (2004) find systematic relationships between ex post mortality and annuity characteristics, such as the timing of payments and the possibility of payments to the annuitant's estate, but they do not find evidence of substantive mortality differences by annuity size. He (2009) finds evidence for the presence of asymmetric information in the life insurance market, a conclusion contrasting with the existing

---

<sup>2</sup> Rothschild and Stiglitz.(1976)

<sup>3</sup> Chiappori (2000) Chiappori and Salanie (2000)

literature. In particular, she finds a significant and positive correlation between the decision to purchase life insurance and subsequent mortality, conditional on risk classification.

One existing explanation for failure to find private information is the presence of heterogeneous preferences for insurance. De Meza and Webb (2001) suggest that there may be advantageous selection where more cautious people are not only more inclined to purchase insurance but also more likely to put effort in preventing risk exposures. The presence of both adverse selection and advantageous selection may create insignificant or even negative correlations between an individual's risk exposure and the decision to purchase insurance even with private information. Finkelstein and McGarry (2006), short F&MG, introduce variables that measure cautiousness and wealth quartiles that are expected to be correlated with insurance demand and risk exposures, and find that these variables positively related to insurance demand and negatively related risk exposures. They explain this result with the existence of multiple types of insurance consumers. More cautious and wealthier individuals are more likely to purchase long-term care insurance and less likely to enter a nursing home. In other words, the presence of asymmetric information is masked by heterogeneous risk attitudes or heterogeneous insurance demand. However, their framework still fails to find private information even after introducing factors about individual heterogeneity, despite a direct evidence of private information when available.

Fang, Keane and Silverman (2008) also provide evidence of advantageous selection. They find that having Medigap insurance would be associated with \$4,000 less in total medical expenditure if not controlling for health status, and \$2,000 more in total medical expenditure if controlling for health status. This result suggests that those who purchase Medigap insurance are healthier, providing evidence of advantageous selection.

This paper makes three contributions to the literature. First, we identify the reasons for failure to detect the private information in the current empirical framework even in the presence of the factors that are related to individuals' heterogeneity of insurance demand. As discussed in F&MG, the presence of individual heterogeneity may cause the problem. However, including

factors that are related to individual heterogeneity may not solve the problem as long as we only have an incomplete set of factors that explain the individual heterogeneity.

Second, we present an alternative test for the asymmetric information in the presence of heterogeneous preferences for insurance. We assume that individuals can be grouped into two types. The *timid* type has a strong taste for insurance and is less likely to experience the insured event. The *bold* type has a weaker preference for insurance and a higher risk of experiencing the insured event. The test of a positive correlation between an individual's risk and the decision to purchase insurance is only valid if conditioning on either type. Since types are not observed, we use the mixture density to jointly model the risk and insurance purchase. Several characteristics that are correlated with the unobserved types are used to probabilistically determine which type the person belongs to. The advantage of this method, however, 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. The intuition of this method is similar to the two-stage instrumental variable model. The factors that explain the individual heterogeneity may be considered as instrumental variables for the unobserved individual heterogeneity. Similar to the two-stage instrumental variable method, the consistency of the second-stage estimator does not require the consistency of the first stages estimates. The literature has been using the mixture density to identify unobserved types. Examples include Lee and Porter (1984), Keane and Wolpin (1997), Knittle and Stengo (2003), and Gan and Mosquera (2008).

Third, we apply this model to the sample of F&MG. We find that the two types of agents behave differently as predicted. The *timid* types are more likely to purchase insurance but less likely to enter into the nursing home than the *bold* types. Conditional on public information and the type of an individual we obtain a statistically significantly positive correlation between ex-post risk and the insurance purchases. This suggests the existence of private information. Allowing for two types of individuals makes it possible to detect the presence of asymmetric

information. We confirm the finding of F&MG without relying on direct evidence of private information. Moreover, we also confirm that the *timid* type would be more likely than the *bold* type to purchase insurance but less likely to use insurance, as suggested by the theory.

Our method is most useful for insurance markets where such direct evidence of private information is not complete or even not available. It makes it possible to account for heterogeneity in risk attitudes and uncover existence of asymmetric information without direct evidence of private information.

The paper is organized as follows, in section 2 we illustrate the identification problem and explain how it is possible to account for heterogeneity in risk preferences. Section 3 describes the data and discusses the results. Section 4 concludes.

## 2. An Empirical Strategy to Detect Private Information in the Insurance Market

We empirically characterize the market for long-term care insurance (*LTCI*) through two equations. The first equation relates individuals' characteristics to the probability of entering a nursing home. The second equation relates the same characteristics to the decision to purchase long-term care insurance.

$$\begin{aligned} NH &= 1(c_\beta + X\beta_X + Z\beta_Z + H\beta_H + u > 0) \\ LTCI &= 1(c_\delta + X\delta_X + Z\delta_Z + H\delta_H + v > 0) \end{aligned} \quad (1)$$

We use  $X$  to denote characteristics that are public information – information that is available to both the individual and the insurance provider. Individuals may have information about the likelihood of eventually entering a nursing home that is not available to insurance providers.

We denote this private information by  $Z$ . Without loss of generality we define  $Z$  so that  $\beta_Z > 0$ .

A higher probability of entering a nursing home creates an incentive to purchase long-term care insurance. As  $Z$  is not reflected in insurance premiums, this leads to a positive relationship between  $Z$  and insurance purchases  $\delta_Z > 0$ . Further, as discussed in the introduction, we include  $H$  to denote individual taste for insurance. We define  $H$  so that  $\delta_H > 0$ , a higher value of  $H$

implies a higher likelihood of purchasing insurance. We assume that the error terms  $u$  and  $v$  are distributed standard normal  $u \sim N(0,1)$  and  $v \sim N(0,1)$ , and are independent on each other,  $Cov(u, v) = 0$ .

If it is not possible to observe  $Z$  or  $H$ , the two equations can be estimated only partially:

$$\begin{aligned} NH &= 1(c_\beta + X\beta_x + u^* > 0) \\ LTCI &= 1(c_\delta + X\delta_x + v^* > 0) \end{aligned} \quad (2)$$

The resulting error terms are given by:  $u^* = H\beta_H + Z\beta_Z + u$  and  $v^* = H\delta_H + Z\delta_Z + v$ .

If individuals have homogeneous risk preferences, there is no variation in  $H$ . The correlation between the error terms is given by:

$$Cov(u^*, v^*) = \rho_1 = \delta'_Z Z' Z \beta_Z > 0, \quad (3)$$

with  $\beta_Z > 0$  and  $\delta_Z > 0$ . Therefore, the presence of unobserved private information leads to a positive correlation, and estimating  $\rho_1$  offers a way to empirically test for the presence of asymmetric information.

However, this test can fail to detect the presence of private information if individuals differ in their inclination to purchase insurance. With heterogeneous risk preferences the correlation of the error terms is described by:

$$\rho_2 = Cov(u^*, v^*) = \delta'_Z Z' Z \beta_Z + \delta'_H H' H \beta_H + \delta'_Z Z' H \beta_H + \delta'_H H' Z \beta_Z \quad (4)$$

The first term in (4),  $\delta'_Z Z' Z \beta_Z$ , is assumed to be positive in the presence of private information. However, the second term  $\delta'_H H' H \beta_H$  is negative if  $\beta_H$  and  $\delta_H$  have opposite signs. In other words, if individuals with a low risk of nursing home use ( $\beta_H < 0$ ) tend to have a taste for insurance ( $\delta_H > 0$ ), the correlation of the two error terms in (4) is no longer indicative of the presence of asymmetric information, but rather a combination of asymmetric information and heterogeneous taste in insurance. Further, the signs of the remaining two terms in (4) cannot be

determined without further assumptions. Without observing  $H$  the sign of  $\rho_2$  in (4) cannot be determined *ex ante*.

In general it is not possible to observe  $H$ . However, it may be possible to observe a set of variables  $W$  that are related to  $H$ . F&MG propose such variables: wealth, adoption of preventive health activities, and seat belt usage. They use these variables to estimate the following model,

$$\begin{aligned} NH &= 1(c_\beta + X\beta_X + W\beta_W + u^* > 0) \\ LTCI &= 1(c_\delta + X\delta_X + W\delta_W + v^* > 0) \end{aligned} \quad (5)$$

If  $W$  fully characterizes  $H$ , the correlation between the error terms can be used to test for private information. However, if  $W$  only represents a subset of variables that characterizes  $H$ , the problem remains. For example, let  $H$  be fully characterized by observed  $W$  and unobserved  $M$ :

$$H = aW + M + \varepsilon, \quad (6)$$

then the error term in the  $NH$  model in (5) is  $u^* = Z\beta_Z + M\beta_H + \varepsilon + u$ , and the error term in the  $LTCI$  model in (5) is  $v^* = Z\delta_Z + M\delta_H + \varepsilon + v$ . The correlation between  $u^*$  and  $v^*$  is given by:

$$\rho_3 = \delta'_Z Z' Z \beta_Z + \delta'_H M' M \beta_H + \delta'_Z Z' M \beta_H + \delta'_H M' Z \beta_Z. \quad (8)$$

Again, similar to  $\rho_2$  in (4), the sign of  $\rho_3$  cannot be determined *ex ante* without further assumptions. The first term in (7),  $\delta'_Z Z' Z \beta_Z$ , is assumed to be positive in the presence of private information, and the second term in (7),  $\delta'_H M' M \beta_H$ , is negative if  $\beta_H$  and  $\delta_H$  have opposite signs. But the signs of the remaining two terms in (7) cannot be determined without further assumptions. Therefore, it is possible that  $\rho_3$  is not positive even if private information is present.

However, in the following discussion, we show that it is possible to solve for this problem if we are willing to assume that the heterogeneity in risk preferences can be captured by allowing each individual to be one of two types with an individual specific probability.

In particular, we assume that there are *bold* ( $B$ ) or *timid* ( $T$ ) individuals.  $H$  takes two values,  $H_B$  and  $H_T$ .<sup>4</sup> Given this assumption, we can rewrite equations in (1) for each type. For the *timid* type individuals ( $H = H_T$ ), we obtain:

$$\begin{aligned}
NH &= 1(c_\beta + X\beta_X + Z\beta_Z + H_T\beta_H + u > 0) \\
&= 1(c_\beta^T + X\beta_X + Z\beta_Z + u > 0) = 1(c_\beta^T + X\beta_X + u^* > 0) \\
LTCl &= 1(c_\delta + X\delta_X + Z\delta_Z + H_T\delta_H + v > 0) \\
&= 1(c_\delta^T + X\delta_X + Z\delta_Z + v > 0) = 1(c_\delta^T + X\beta_X + v^* > 0)
\end{aligned} \tag{9}$$

In both equations in (9), the effect of  $H$  is absorbed into the constant terms,  $c_\beta^T$  and  $c_\delta^T$ .

Similarly, for the *bold* type individuals ( $H = H_B$ ), we obtain:

$$\begin{aligned}
NH &= 1(c_\beta^B + X\beta_X + u^* > 0) \\
LTCl &= 1(c_\delta^B + X\delta_X + v^* > 0)
\end{aligned} \tag{10}$$

Again, the constant terms  $c_\beta^B$  and  $c_\delta^B$  absorb the effect of  $H$  while the error terms include the private information. The model predicts the relative magnitude of the constant terms. Everything else equal, a *timid* type individual would be more likely to purchase  $LTCl$  but less likely to enter the nursing home than a *bold* type individual, i.e.  $c_\delta^T > c_\delta^B$  and  $c_\beta^T < c_\beta^B$ , everything else being equal. More importantly, the error terms  $u^*$  and  $v^*$  now only include the private information  $\mathbf{Z}$ , but not the individual heterogeneity. Therefore, the correlation between  $u^*$  and  $v^*$  reflects the presence of private information  $\mathbf{Z}$ . Imposing the two type structure transforms the problem from one of identifying  $H\beta_H$  and  $H\delta_H$  in (1), to a problem of identifying  $c_\beta^{T(B)}$  and  $c_\delta^{T(B)}$ , and – at least probabilistically – the type of an individual.

We need to jointly identify  $c_\beta^T$ ,  $c_\delta^T$ ,  $c_\beta^B$ ,  $c_\delta^B$ ,  $\beta_X$ ,  $\delta_X$ , the probability of belonging to a certain type, and the correlations between the two error terms  $u^*$  and  $v^*$ . Conditional on  $X$  and

---

<sup>4</sup> More generally, both  $H_B$  and  $H_T$  can be random variables that are uncorrelated with  $X$  and  $Z$ .

$W$ , we observe four possible outcomes,  $(NH=i, LTCI=j)$ , for  $i = B, T$  and  $j = B, T$ . The probability of a given outcome depends on the type of the individual:

$$\begin{aligned} & \Pr(NH = i, LTCI = j) \\ &= \Pr(NH = i, LTCI = j | H = H_T) \Pr(H = H_T) \\ & \quad + \Pr(NH = i, LTCI = j | H = H_B) \Pr(H = H_B) \end{aligned} \quad (11)$$

As described in equation (6),  $H$  is determined by  $W$  and  $M$ . We assume that the probability of being of a certain type varies with  $W$ :

$$\Pr(H = H_T | W, M) \neq \Pr(H = H_T), \quad \text{for some } W \text{ and } M. \quad (12)$$

If we assume that  $W$  is observed but  $M$  is not observed, it is not possible to consistently estimate the coefficient  $\gamma$  in the equation below:

$$\Pr(H = H_T | W, M) = \Pr(W\gamma + M + \varepsilon > 0) \quad (13)$$

Assuming that the unobserved  $M$  can be written as a linear function of  $W$  and an error term, i.e.,  $M = W\alpha + \tau$ ; equation (13) can be rewritten as:

$$\Pr(H = H_T | W) = \Pr(W\gamma + W\alpha + \tau + \varepsilon > 0) = \Pr(W\gamma^* + \omega > 0) \quad (14)$$

where the random errors  $\tau$  and  $\varepsilon$  are normally distributed, and the parameter  $\gamma^*$  is the sum of  $\gamma$  and  $\alpha$ , scaled by a constant such that  $\omega \sim N(0, 1)$ . In the linear model,  $\gamma$  may be consistently estimated if  $M$  is uncorrelated with  $W$  (where  $\alpha = 0$ ). However, in the nonlinear setting here,  $\gamma$  cannot be consistently estimated regardless of the correlation between  $M$  and  $W$ .

The key identifying assumption is that – conditional on the type of an individual –  $W$  and  $M$  are not related to either the probability of entering a nursing home, or the probability to purchase long-term care insurance:

$$\Pr(NH = i, LTCI = j | H = H_T, W, M) = \Pr(NH = i, LTCI = j | H = H_T) \quad (15)$$

Consequently, any association between  $W$  and  $M$  and the probability of entering a nursing home or purchasing insurance is solely driven by the association between  $W$  and  $M$  and the probability to belong to a certain type.

Intuitively, this identification assumption is similar to the identification assumption of the instrumental variable model.  $\mathbf{W}$  and  $\mathbf{M}$  may be considered as the “instrumental variables” for the type variable  $H$ . They are assumed to be uncorrelated with  $NH$  and  $LTCI$  but correlated with  $H$ . More generally,

$$\begin{aligned} & \Pr(NH = i, LTCI = j | \mathbf{W}, \mathbf{M}) \\ &= \Pr(NH = i, LTCI = j | H = H_T) \Pr(H = H_T | \mathbf{W}, \mathbf{M}) \\ & \quad + \Pr(NH = i, LTCI = j | H = H_B) \Pr(H = H_B | \mathbf{W}, \mathbf{M}) \end{aligned} \quad (16)$$

Again, the relative contribution of  $\mathbf{W}$  and  $\mathbf{M}$  to the variation in  $H$  does not affect

$\Pr(NH = i, LTCI = j)$  conditional on the type of an individual. Rewriting (16) conditional on the observed variables  $\mathbf{X}$  and  $\mathbf{W}$  gives:

$$\begin{aligned} & \Pr(NH = i, LTCI = j | \mathbf{X}, \mathbf{W}) \\ &= \Pr(NH = i, LTCI = j | H = H_T, \mathbf{X}) \Pr(H = H_T | \mathbf{W}) \\ & \quad + \Pr(NH = i, LTCI = j | H = H_B, \mathbf{X}) \Pr(H = H_B | \mathbf{W}) \end{aligned} \quad (17)$$

All terms in (17) are previous discussed in (9), (10), and (14). Therefore, one may construct a likelihood function to estimate such a model.

**Lemma:** If the probability of being of a certain type varies with  $\mathbf{W}$  (equation (12)), and conditional on the type of an individual  $\mathbf{W}$  is not related to either the probability of entering a nursing home or the probability to purchase long-term care insurance (equation (15)), then estimating equations (9) and (10) together with either (13) or (14) produces consistent estimates of the parameters  $c_\beta^T, c_\delta^T, c_\beta^B, c_\delta^B, \beta_X, \delta_X$ , and the correlations between the two error terms  $u^*$  and  $v^*$ .

Equations (16) and (17) reveal that (13) and (14) result in different estimates for the coefficient of  $\mathbf{W}$ , but the other coefficients of the model are not affected by the fact that  $\mathbf{M}$  is not observed.

If we have more than one dimension of information in  $\mathbf{W}$ , we may only use a subset of  $\mathbf{W}$  to estimate the model. One implication of this model is that the parameters of interest,  $c_\beta^T, c_\delta^T$ ,

$c_{\beta}^B$ ,  $c_{\delta}^B$ ,  $\beta_X$ ,  $\delta_X$  and the correlation between  $u^*$  and  $v^*$  remain to be consistently estimated while coefficients of  $W$  will not.

As briefly discussed earlier, the intuition of the method is very similar to the two-stage instrumental variable model. The consistency of the 2SLS estimates does not require the consistency of the first-stage regression. Similarly, the fact that we do not observe  $M$  does not create inconsistent estimates of the parameters of interest. Therefore, the advantage of the proposed method is that it only requires some (but not full) information about  $H$  to identify the parameters of interest. In comparison, the method in the literature, as used by F&MG, replaces the unobserved type variable  $H$  by a set of proxies  $W$  in the  $NH$  and  $LTCI$  equations. It works only if  $W$  completely characterizes  $H$ . However, given how little we understand the unobserved heterogeneity  $H$ , this is unlikely to hold in practice.

### 3. Data and Results

We illustrate our estimation procedure by applying it to the data assembled by F&MG. The data are based on the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). This survey is designed to be representative of the non-institutionalized US population born in 1923 or earlier and their spouses. For a more detailed discussion see F&MG.

It is possible to observe insurance status, nursing home utilization, and a number of demographic and health variables that make it possible to control for risk classification of individuals by insurers. 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,  $X$ .

The data also contain information that is not used by insurers to set premiums. 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 self-reported probability of entering nursing home has been shown to be consistent on average with observed probabilities at the aggregate level, but has serious reporting errors at individual level (see, for example, Hurd and McGarry 2002; Gan, Hurd and McFadden 2004), suggesting that the measure can best serve a noisy proxy to the private information.

The data also contain information about wealth and proxy 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.

Table 1 displays the descriptive statistics. The sample contains 5,119 individuals. 10% of them have long-term care insurance in 1995 and 16% enter a nursing home at some point from 1995-2000. However, to be comparable across various specifications, we limit our sample to individuals without missing information for any utilized variable. The working sample size consists of 5,000 observations.

We first estimate model (2) for only one type of agent to provide a baseline for our further analysis. In all specifications, we control in both equations for the public information available to the insurance company,  $X$ , summarized by the company predictor variable.

The estimates reported here are similar to those reported by F&MG.<sup>5</sup> In the first column in Table 2, we confirm that the company prediction has a positive effect on the probability of entering a nursing home. The estimated coefficient for the company predictor ( $\beta_X$ ) is 1.805 (0.909).<sup>6</sup> However, the insurance company predictor reduces the probability of purchasing insurance. The estimated coefficient ( $\delta_X$ ) is -0.694 (0.123), corresponding a marginal effect of -0.129. We obtain a negative (not significantly different from zero) estimate

---

<sup>5</sup> To make our estimates comparable to each other our specification differs slightly from those in F&MG. When using the exact specifications as that of F&MG, we obtain the same results.

<sup>6</sup> This corresponds to a marginal effect of 0.40 if other variables are evaluated at their means (See table A1).

at -0.036 (0.041) for the correlation between the two error terms. In other words the correlation test does not provide evidence for the existence of asymmetric information.

Next, in the second column in Table 2, we add proxies for private information, individual predictions to enter nursing homes,  $Z$ , to the two equations. The coefficient for this variable is positive in both equations, suggesting that indeed private information is present.

The third column in Table 2 displays the results for the model after we added proxies for the risk attitudes (types) of individuals,  $W$ , but without the individual prediction. These proxy variables for risk attitudes are dummy variables for “preventative health measures taken”, seat belt usage and for the 4<sup>th</sup>, 3<sup>rd</sup>, and 2<sup>nd</sup> wealth quartile. We confirm that the coefficients for the variables in  $W$  have opposite signs in the two equations. F&MG suggest that different signs in  $W$  in two equations actually indicate the heterogeneity in tastes.

Now, we estimate the model with two types of individuals. We jointly estimate (14), (9), and (10). We let  $H = 1$  be the *timid* type and  $H = 0$  be the *bold* type.  $W$  consists of seat belt usage, preventative healthcare measures, and wealth quartiles.

We restrict the coefficients  $\beta_X$ ,  $\delta_X$  and the correlation  $\rho$  to be identical for the two types, but allow the constant term differ. Column 1 in Table 3 displays the results (the corresponding marginal effects are shown in Table A2). The top panel illustrates the effect 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. A person who takes preventative activities has a 10 percentage points higher probability to be of the timid type. Always wearing seat belt increases this probability by 12 percentage points. Finally, compared with individuals in the fourth (lowest) wealth quartile, having a wealth level in the top quartile (first quartile) increases the probability of belonging to the timid type by 31 percentage points, for individuals in the second quartile the increase is 21 percentage points, and for individuals in the third quartile it is 12 percentage points.

People who belong to different types exhibit clear differences in their behavior. As predicted by the model, we find that a *timid*-type individual is more likely to purchase the *LTCI* but the less likely to enter into a nursing home. 98.3% of those purchasing *LTCI* are categorized as the *timid* type, while 13.27% of those who enter into a nursing home are categorized as the *timid* type. The lower panel of Table 3 displays the relationship between individual characteristics and insurance purchases and nursing home usage.

As discussed in the previous subsection, the proposed model has clear predictions on the relative magnitude of the constant terms. For the *LTCI* model, the estimated constant  $c_{\delta}^T$  for the *timid* type is -0.269 (0.188), significantly larger than the estimated constant  $c_{\delta}^B$  for the *bold* type at -2.312 (0.237). A one-sided Z-test rejects the null hypothesis that  $\hat{c}_{\delta}^T \leq \hat{c}_{\delta}^B$  ( $p$ -value = 0.000). For the *NH* model, the estimated constant for the *timid* type  $c_{\beta}^T$  is -2.288 (0.197), statistically smaller than the estimated constant for the *bold* type  $c_{\beta}^B$  at -1.269 (0.061),  $p$ -value = 0.000. Both test results are consistent with the predictions of the model. In terms of probabilities of purchasing *LTCI* and entering nursing homes, the *timid* type would be 35.98 percentage points more likely to purchase *LTCI* but 15.78 percentage points less likely to enter into nursing homes.

Most importantly, by separating individuals into two types, we are able to obtain clear evidence of private information. The estimated correlation between the error terms in the *NH* model and the *LTCI* model is positive and statistically significant at 0.621 (0.263). It is important to note that this is achieved without using any data on private information.

The second set of estimates in Table 3 includes one dimension of private information, the individual prediction of entering nursing homes, in both the *LTCI* equation and the *NH* equation. As in the case of one type model in Table 2, the coefficient of this variable is positive in both *LTCI* and *NH* equations, showing the importance of such private information in determining both

the decisions to buy *LTCI* and to enter nursing homes. Adding the proxy for private information reduces the correlation between the two error terms to 0.566 (0.209), although it remains positive and statistically significant. This result suggests that (a) the individual prediction of entering nursing home as elicited in the survey may only characterize a small portion of the private information; and (b) adding more and more private information may eventually lead to zero correlation, as predicted by the model.

Unlike the one-type model, our two-type model can detect the existence of the private information even in the absence of observable data on private information. If some data on private information is observed our model can indicate to what extent additional unobserved private information influences the decision to purchase insurance.

As discussed in the previous subsection, one implication of our model is that even a partial set of  $W$  may produce consistent estimates of the key parameters of interest. Therefore, as a further test of the model, we vary the choice of variables in determining types,  $W$ . In our current setting, the set  $W$  consists of preventive care, always wearing seat belts, and wealth quartiles. As discussed in F&MG, the first two variables describe “cautiousness” while the wealth quartiles could be sources of heterogeneous private insurance demand.

In the first two columns in Table 4, we do not include any variables in  $W$ . In column (1) we only include the company predictor,  $X$ , while column (2) includes individual private information on entering nursing homes. The insurance company predictor remains positive in the *NH* equation and negative in the *LTCI* equation. More importantly, the constant in both equations satisfy the predictions of our two-type model. However, the correlation between the error terms in both *NH* equation and the *LTCI* equation is no longer significant. This result is expected since without the any information to economically distinguish the two types of consumers, the type is purely identified by functional form. Therefore, one of our two identification assumptions – the probabilities of purchasing insurance and nursing home use have to vary with  $W$  – is violated.

The second column includes private information on entering nursing home. The coefficient estimates for the private information variable are positive and significant in both *NH* and *LTCI* equations. Again, the correlation between the error terms is not statistically significant.

The third and fourth column in Table 4 show the results when the wealth quartiles information is used. Our model suggests that parameter estimates in both *NH* and *LTCI* equations should be similar to the corresponding parameter estimates in Table 3 when all available information in *W* is used. Both sets of estimates are indeed similar to each other. For example, in the column (3) in Table 4, the coefficient estimate for the insurance company predictor is 1.828 (0.104) in the *NH* equation. The corresponding coefficient estimate in Table 3 is 1.833 (0.101). The coefficient estimate for constant term in *NH* (*timid* type) equation in Table 4 is -2.203 (0.237) while the corresponding coefficient estimate in Table 3 is -2.288 (0.237). The correlation between the error terms in the *NH* model and the *LTCI* model is positive and statistically significant. It is 0.595 (0.266) without any direct information on private information and – as expected – drops to .472(275) after adding private information in column (4).

The fifth and sixth column in Table 4 list results only using variables of preventive activities with the third column only has the company predictor while the fourth column has both the company predictor and private predictor. The coefficient estimates in both *NH* equation and *LTCI* equation have the expected signs and are statistically significant. The estimates of constants are also consistent with the predictions of our model.

Table 5 presents a formal test. The test compares the estimates of the parameters of interest in the *NH* and *LTCI* equations presented in Table 3 and Table 4. Our theoretical analysis suggests that having a partial set of *W* may affect the estimates for the coefficient estimates for *W* in the type equation but not the coefficients in the *NH* and *LTCI* equations. The second panel in the table compares estimates from the full model with the model to the model with only wealth quartiles. As predicted, the estimates in the *NH* and *LTCI* equations in two models are not statistically different with each other. Interestingly, the estimates in the type-determination

equation from these two sets of estimates are not different from each other, either. One potential reason for this may be due to fact that the omitted sets of  $W$  are independent with each other. Finally, the first panel in the table compares the full model with the model with only prevention variables as  $W$ . Testing results indicate that estimates from the two models are statistically different from each other for both the coefficients in the *LTCI* and *NH* equations and for the type equations. One possible reason is that preventive variables may not be sufficient to identify the model, similar to the weak instrumental variable problem in the literature. The true reason remains to be understood.

## 5. Conclusions

Identifying private information in the insurance markets is important for empirically test the economic theories of moral hazard and adverse selection. It is also useful to improve efficiency of the insurance market. This paper proposes and estimates a new method to identify private information in the presence of heterogeneity consumer types. We illustrate this method for an insurance market where direct evidence for private information is available. We are able to detect the presence of asymmetric information without using this direct evidence.

In particular, this paper makes three contributions to the literature. First, when only a partial set of information is available to characterize the individual heterogeneity, this paper suggests that the current method of testing positive relations between the demand of insurance model and the usage of the service model to identify private information is no longer applicable. Second, based on this discussion, this paper provides a new method to identify private information with individuals' risk heterogeneity and only a partial set of information to characterize such risk heterogeneity. This method assumes that individuals' risk heterogeneity can be grouped into two unobserved categories. Third, although private information is known to be present in the long term care market, it cannot be detected by the existing method but is detected by the proposed method.

The identification of this method is similar to that of the instrumental variable model. It requires the variables characterizing risk heterogeneity are uncorrelated with decisions to purchase insurance and to use the service covered by the insurance. The procedure described here is a tool that can be used to detect the presence of private information in insurance markets with heterogeneous risk preferences when direct evidence of private information is not available.

## References

- Cawley, John and Tomas Philipson (1999). "An empirical examination of information barriers to trade in insurance." *American Economic Review*, 1999, 89 (4), pp.827-846.
- Chiappori, Pierre-Andre (2000). "Econometric models of insurance under asymmetric Information." In Georges Dionne, ed. *Handbook of Insurance Economics*. London: Kluwer.
- Chiappori, Pierre-Andre and Bernard Salanie (2000). "Testing for asymmetric information in insurance Markets." *Journal of Political Economy*, 2000, 108 (1), pp. 56-78.
- Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie (2006). "Asymmetric information in insurance: general testable implications." *Rand Journal of Economics*.
- Cohen, Alma (2005). "Asymmetric information and learning: evidence from the automobile insurance." *The Review of Economics and Statistics*. May 2005 82(2): 197-207.
- de Meza, David and Webb, David C. (2001) "Advantageous selection in insurance markets." *Rand Journal of Economics*, Summer 2001 32 (2), pp. 249-262
- Finkelstein Amy, and Kathleen McGarry (2006). "Multiple dimensions of private information: evidence from the long-term care insurance market." *American Economic Review*, Vol. 96 No.4, Page 938-958.
- Finkelstein, Amy and James Poterba (2002). "Selection effects in the market for individual annuities: new evidence from the United Kingdom." *Economic Journal*, 2002, 112 (476), pp. 28-50.
- Hanming Fang, Michael P. Keane, and Dan Silverman (2008). "Sources of advantageous selection: evidence from the Medigap insurance market." *Journal of Political Economy*. Volume 116, Issue 2, Page 303–350, Apr 2008
- He, Daifeng (2009). "The life insurance market: asymmetric information revisited." *Journal of Public Economics*.
- Gan Li, and Roberto Mosquera (2008). "An empirical study of the credit market with unobserved consumer types." NBER WPS #13873.
- Lee, Lung-Fei, and Robert Porter (1984). "Switching regression models with imperfect sample separation information: application on cartel stability." *Econometrica*, Vol. 52, No. 2 (Mar., 1984), pp. 391-418
- Rothschild, Michael and Joseph Stiglitz. "An essay on the economics of imperfect information." *Quarterly Journal of Economics* 1976, 90 (4) pp. 629-649.

Smart, Michael. "Competitive insurance markets with two unobservables (2004)." *International Economic Review*, 2000, 41(1) pp. 153-169.

**Table 1: Descriptive Statistics**

Variable	Mean	sd	Min	Max
Nursing Home Use	0.163	0.369	0	1
Long Term Care Insurance	0.108	0.311	0	1
Insurance company prediction	0.218	0.231	0.006	1
Individual prediction	0.177	0.248	0	1
Preventive health activity	0.659	0.304	0	1
Always wear seat belt	0.768	0.422	0	1
Top quartile of wealth	0.285	0.451	0	1
Wealth quartile 2	0.270	0.444	0	1
Wealth quartile 3	0.243	0.429	0	1

Note: The sample consists of the elderly aged 78 on average in 1995 who reported long-term care insurance status and nursing home use from 1995 to 2000 from the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (5,000 observations).

**Table 2: One type Bivariate Probit Model**

		Company prediction	Company and individual prediction	Company prediction, prevention and wealth quartiles	Company and individual prediction, prevention and wealth quartiles
		(1)	(2)	(3)	(4)
<i>NH</i>	Insurance company prediction	1.805*** (0.090)	1.786*** (0.091)	1.708*** (0.093)	1.683*** (0.094)
	Individual prediction		0.186** (0.092)		0.208** (0.092)
	Preventive health activity			-0.176** (0.081)	-0.187** (0.081)
	Always wear seat belt			-0.114** (0.056)	-0.116** (0.056)
	Top quartile of assets			-0.125* (0.072)	-0.124* (0.072)
	Wealth quartile 2			-0.070 (0.071)	-0.071 (0.071)
	Wealth quartile 3			0.026 (0.071)	0.026 (0.071)
	Constant	-1.459*** (0.034)	-1.490*** (0.037)	-1.188*** (0.081)	-1.213*** (0.082)
	<i>LTCI</i>	Insurance company prediction	-0.694*** (0.123)	-0.781*** (0.126)	-0.431*** (0.128)
Individual prediction			0.547*** (0.094)		0.538*** (0.097)
Preventive health activity				0.162* (0.095)	0.134 (0.096)
Always wear seat belt				0.234*** (0.068)	0.232*** (0.068)
Top quartile of assets				0.592*** (0.088)	0.596*** (0.089)
Wealth quartile 2				0.424*** (0.090)	0.421*** (0.091)
Wealth quartile 3				0.275*** (0.093)	0.272*** (0.094)
Constant		-1.092*** (0.034)	-1.184*** (0.038)	-1.836*** (0.112)	-1.904*** (0.116)
Correlation of two error terms $\rho$		-0.036 (0.041)	-0.044 (0.040)	-0.015 (0.041)	-0.023 (0.041)
Number of observations	5,000	5,000	5,000	5,000	
Log-Likelihood	-3713.25	-3698.15	-3657.14	-3642.50	

Notes: Estimation of a bivariate probit of any nursing home use (1995-2000) and long-term care insurance coverage (1995). \*\*\*, \*\*, \* denote statistical significance at the 1%, 5%, and 10% level, respectively. Our estimates are weighted using the 1995 household weights.

**Table 3: Two-type Model**

	Company predictor only (1)		Company and Individual predictors (2)	
<b>Type (<i>timid type</i> = 1)</b>				
Preventive health activity	0.308**		0.273*	
	(0.145)		(0.151)	
Always wear seat belt	0.382***		0.393***	
	(0.104)		(0.104)	
Top quartile of assets	0.852***		0.876***	
	(0.132)		(0.133)	
Wealth quartile 2	0.593***		0.605***	
	(0.133)		(0.133)	
Wealth quartile 3	0.349***		0.355***	
	(0.133)		(0.134)	
Constant	-1.574***		-1.546***	
	(0.185)		(0.183)	
	<i>Timid type</i>	<i>Bold type</i>	<i>Timid type</i>	<i>Bold type</i>
<b>NH</b>				
Insurance company prediction	1.828***	1.828***	1.798***	1.798***
	(0.104)	(0.104)	(0.103)	(0.103)
Individual prediction			0.211**	0.211**
			(0.098)	(0.098)
Constant	<b>-2.288***</b>	<b>-1.269***</b>	<b>-2.254***</b>	<b>-1.303***</b>
	(0.215)	(0.061)	(0.239)	(0.063)
<b>LTCI</b>				
Insurance company prediction	-0.628***	-0.628***	-0.751***	-0.751***
	(0.184)	(0.184)	(0.190)	(0.190)
Individual prediction			0.824***	0.824***
			(0.197)	(0.197)
Constant	<b>-0.269</b>	<b>-2.312***</b>	<b>-0.426**</b>	<b>-2.498***</b>
	(0.188)	(0.237)	(0.166)	(0.331)
<b><math>\rho</math></b>	0.621**		0.566***	
	(0.271)		(0.209)	
Loglikelihood	-3658.49		-3643.55	
Number of Obs	5,000		5,000	

Notes: Estimation of a two-type model estimation of any nursing home use (1995-2000) and long-term care insurance coverage (1995).

\*\*\*, \*\*, \* denote statistical significance at the 1%, 5%, and 10% level, respectively.

Our estimates are weighted using the 1995 household weights.

Table 4: Two-type Model: Robustness check

	Constant		Wealth quartiles		Preventive activity and Seat belt	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Pr(<i>timid</i> type = 1)</b>						
Preventive health activity					3.379** (1.676)	3.306** (1.636)
Always wear seat belt					2.892** (1.331)	2.936** (1.390)
Top quartile of wealth			0.913*** (0.136)	0.944*** (0.140)		
Wealth quartile 2			0.643*** (0.129)	0.657*** (0.132)		
Wealth quartile 3			0.387*** (0.124)	0.391*** (0.126)		
Constant	0.459*** (0.052)	0.463*** (0.051)	-1.234*** (0.189)	-1.188*** (0.178)	-4.316** (1.871)	-4.291** (1.898)
<b>NH (<i>timid</i> type)</b>						
Insurance company prediction	8.763*** (1.462)	8.702*** (1.409)	1.833*** (0.101)	1.799*** (0.105)	1.756*** (0.091)	1.733*** (0.092)
Individual prediction		0.377* (0.209)		0.199** (0.097)		0.206** (0.092)
Constant $c_{\beta}^T$	<b>-8.463***</b> (1.301)	<b>-8.463***</b> (1.255)	<b>-2.203***</b> (0.237)	<b>-2.120***</b> (0.396)	<b>-1.544***</b> (0.057)	<b>-1.581***</b> (0.061)
<b>NH (<i>bold</i> type)</b>						
Insurance company prediction	8.763*** (1.462)	8.702*** (1.409)	1.833*** (0.101)	1.799*** (0.105)	1.756*** (0.091)	1.733*** (0.092)
Individual prediction		0.377* (0.209)		0.199** (0.097)		0.206** (0.092)
Constant $c_{\beta}^B$	<b>-1.370***</b> (0.123)	<b>-1.424***</b> (0.129)	<b>-1.313***</b> (0.061)	<b>-1.343***</b> (0.066)	<b>-1.329***</b> (0.053)	<b>-1.357***</b> (0.055)
<b>LTCI (<i>timid</i> type)</b>						
Insurance company prediction	-0.696*** (0.124)	-0.783*** (0.127)	-0.763*** (0.213)	-0.873*** (0.215)	-0.593*** (0.125)	-0.684*** (0.128)
Individual prediction		0.547*** (0.094)		0.923*** (0.275)		0.525*** (0.096)
Constant $c_{\delta}^T$	<b>-1.100***</b> (0.057)	<b>-1.185***</b> (0.058)	<b>-0.089</b> (0.273)	<b>-0.309</b> (0.209)	<b>-0.963***</b> (0.059)	<b>-1.057***</b> (0.059)
<b>LTCI (<i>bold</i> type)</b>						
Insurance company prediction	-0.696*** (0.124)	-0.783*** (0.127)	-0.763*** (0.213)	-0.873*** (0.215)	-0.593*** (0.125)	-0.684*** (0.128)
Individual prediction		0.547*** (0.094)		0.923*** (0.275)		0.525*** (0.096)
Constant $c_{\delta}^B$	<b>-1.074***</b> (0.098)	<b>-1.179***</b> (0.100)	<b>-2.317***</b> (0.256)	<b>-2.564***</b> (0.591)	<b>-1.387***</b> (0.080)	<b>-1.464***</b> (0.082)
$\rho$	<b>-0.097</b> (0.114)	<b>-0.096</b> (0.113)	<b>0.595**</b> (0.266)	<b>0.472*</b> (0.275)	<b>-0.017</b> (0.042)	<b>-0.025</b> (0.042)
Loglikelihood	-3677.7008	-3663.555	-3673.8107	-3658.1565	-3683.5369	-3669.3261
Number of Obs	5,000	5,000	5,000	5,000	5,000	5,000

Notes: The same as Table 4.

Table 5: Hausman test: Baseline model vs Robust check

	Baseline model vs wealth only		Baseline model vs Preventive activity & Seat belt only	
	Company prediction	Both predictions	Company prediction	Both predictions
Type equation	1.824 (0.768)	2.162 (0.706)	9.185 (0.027)	8.999 (0.029)
<i>NH</i> and <i>LTCI</i> equations	1.068 (0.998)	1.033 (0.999)	38.302 (0.000)	34.243 (0.000)

Notes: Table reports Hausman test statistics and p-values in the parenthesis.

**Table A1: One type Bivariate Probit Model: Marginal effects at average**

	Company prediction (1)	Company and Individual prediction (2)	Company prediction and wealth (3)	Company and Individual prediction and wealth (4)
<i>NH</i>				
Insurance company prediction	0.400*** (0.021)	0.396*** (0.021)	0.376*** (0.021)	0.370*** (0.021)
Individual prediction		0.041** (0.020)		0.046** (0.020)
Preventive health activity			-0.039** (0.018)	-0.041** (0.018)
Always wear seat belt			-0.025** (0.012)	-0.026** (0.012)
Top quartile of assets			-0.028* (0.016)	-0.027* (0.016)
Wealth quartile 2			-0.015 (0.016)	-0.016 (0.016)
Wealth quartile 3			0.006 (0.016)	0.006 (0.016)
<i>LTCI</i>				
Insurance company prediction	-0.129*** (0.023)	-0.143*** (0.023)	-0.076*** (0.023)	-0.091*** (0.023)
Individual prediction		0.100*** (0.017)		0.094*** (0.017)
Preventive health activity			0.029* (0.017)	0.023 (0.017)
Always wear seat belt			0.041*** (0.012)	0.041*** (0.012)
Top quartile of assets			0.105*** (0.015)	0.104*** (0.015)
Wealth quartile 2			0.075*** (0.016)	0.074*** (0.016)
Wealth quartile 3			0.049*** (0.016)	0.048*** (0.016)
Number of observations	5,000	5,000	5,000	5,000

Notes: Table reports marginal effects of marginal success probability of entering nursing home and long-term care insurance coverage from bivariate probit estimation of equation (2) and (5).

\*\*\*, \*\*, \* denote statistical significance at the 1%, 5%, and 10% level, respectively.

Our estimates are weighted using the 1995 household weights.

**Table A2: Two-type Model: Marginal Effects**

	Company predictor only		Company & Individual predictors	
<b>Type (<i>timid</i> type = 1)</b>				
Preventive health activity	0.106*		0.097	
	(0.062)		(0.061)	
Always wear seat belt	0.127**		0.131**	
	(0.056)		(0.057)	
Top quartile of assets	0.276***		0.286***	
	(0.099)		(0.100)	
Wealth quartile 2	0.195**		0.201**	
	(0.080)		(0.080)	
Wealth quartile 3	0.118*		0.121*	
	(0.062)		(0.062)	
	<i>Timid type</i>	<i>Bold type</i>	<i>Timid type</i>	<i>Bold type</i>
<b><i>NH</i></b>				
Insurance company prediction	0.555*	1.273***	0.560*	1.237***
	(0.307)	(0.206)	(0.306)	(0.207)
Individual prediction			0.078	0.132***
			(0.068)	(0.042)
<b><i>LTCI</i></b>				
Insurance company prediction	-0.159***	-0.470*	-0.200***	-0.552*
	(0.040)	(0.216)	(0.053)	(0.239)
Individual prediction			0.278	0.544***
			(0.196)	(0.059)
Number of Obs	5,000		5,000	

Notes: Table reports marginal effects of marginal success probability of entering nursing home and long-term care insurance coverage from two-type model of equation (15). For the type equation, we estimate the marginal success probability of being timid type. \*\*\*, \*\*, \* denote statistical significance at the 1%, 5%, and 10% level, respectively. Standard errors are calculated by Monte Carlo simulation.