The Asymmetric Information Problem in Taiwan's Cancer Insurance Market

Kili C. Wang[a,b], Jin-Lung Peng[c], Yi-Yun Sun[d] and Yao-Chia Chang[e]

[a]Department of Insurance, Tamkang University, Taipei, Taiwan.
E-mail: kili@mail.tku.edu.tw
[b]Risk and Insurance Research Center, College of Commerce, National Chengchi University, Taipei, Taiwan.
[c]Department of Risk Management and Insurance, National Chengchi University, Taipei, Taiwan.
[d]Graduate Institute of Finance and Insurance, Shih Chien University, Taipei, Taiwan.
[e]Graduate Institute of Finance and Insurance, Tamkang University, Taipei, Taiwan.

This paper investigates the problem of asymmetric information in Taiwan's cancer insurance market. Through the survey data, we find evidence of adverse selection existing in this market. Furthermore, we collect additional information on the individual, and find that the individual's family cancer history contains additional valuable information. It can not only more accurately predict the probability of contracting cancer, as well as predict the willingness to purchase extended cancer insurance, but it can also help to mitigate the severity of adverse selection in the insurance market.

Introduction

This paper intends to test whether asymmetric information problems exist in Taiwan's cancer insurance market. As far as we know, we are the first to investigate the asymmetric information problems in the cancer insurance market. Although cancer has been the most severe mortal disease in the world and the cancer insurance market is nowadays becoming increasingly important; the asymmetric information problems in this market have received less attention than those in other insurance markets.[1] This paper intends to fill this gap in the literature.

---

[1] For instance, Puelz and Snow (1994), Richaudeau (1999), Chiappori and Salanie (2000), Dionne *et al.* (2001), Cohen (2005), Saito (2006) and Cohen and Einav (2007) have focused on the automobile insurance market, Cardon and Hendel (2001) on the health insurance market,

We start by testing the conditional relationship between the occurrence of the risk and the choice of the coverage. Although Taiwan's compulsory National Health Insurance (NHI) has provided basic coverage for cancer, many people still purchase extended cancer insurance from commercial insurance companies.[2] Thus, we investigate the existence of asymmetric information by examining whether an individual who voluntarily purchases extended cancer insurance has a greater chance of suffering from cancer. On the basis of the theory of adverse selection or moral hazard,[3] those who purchase extended cancer insurance contracts are the ones with a higher probability of contracting cancer. On the other hand, according to the theory of advantageous selection,[4] those who purchase extended cancer insurance contracts are the ones with a lower probability of contracting cancer.

When testing the conditional correlation between coverage and risk in Taiwan's cancer insurance market, we adopt two different methodologies. One involves pairwise conditional correlation analysis which is based on the methodology of Chiappori and Salanie (2000), while the other is the two-stage method which is similar to the methodology of Dionne *et al.* (2001). These two methodologies serve as robustness tests for each other.

Our empirical evidence strongly supports the existence of a positive relationship between risk and choice in the cancer insurance market. When the positive conditional correlation is identified, it is usually not easy to distinguish adverse selection from moral hazard. Some studies have tried to disentangle them from each other by quasi natural experiments,[5] dynamic

---

Finkelstein and Poterba (2004) on the annuities market, Cawley and Philipson (1999) on the life insurance market, and Finkelstein and McGarry (2006) on the long-term care insurance market.

[2] Such extended cancer insurance offers additional coverage that includes new medicine and new technologies for curing the disease, the fees for hospital beds beyond what the NHI offers, disability-related income subsidies, and long-term health-care subsidies.

[3] On adverse selection, Rothschild and Stiglitz (1976) argued that the high risk insured chooses a high coverage contract, and the low risk insured chooses low risk coverage through self-selection in a world with asymmetric information. On moral hazard, Shavell (1979) inferred that the insured who purchased higher (lower) coverage would pay less (more) attention to taking care, and that the probability of an accident occurring would be higher (lower).

[4] Some studies have inferred that some individuals are more willing to both purchase insurance and take precautions to prevent accidents. Hence, the individual who purchases more coverage may also have a lower *ex post* risk probability. The literature has explained the motivation behind the individual's precautionary effort in relation to risk aversion (de Meza and Webb, 2001), risk perception (Koufopoulos, 2003; Huang, Liu and Tzeng, 2006), patience (Sonnenholzner and Wambach, 2006), and regret (Huang *et al.*, 2008).

[5] For example, in Chiappori and Salanie (2000), the young drivers' high coverage contracts choices are affected by their parents' good past records, and this could induce the problem of moral hazard. In Cardon and Hendel (2001), the employee's coverage choice is exogenously made by the employer, and this also could induce the problem of moral hazard.

data,[6] or the natural characteristics of some special insurance markets.[7] This paper benefits from a characteristic of the cancer insurance market, that is, the people who purchase extended cancer insurance might not really reduce their efforts to prevent themselves from suffering from cancer. Hence, we could assume that there is very little chance that moral hazard plays a role in the cancer insurance market.[8] Under such an assumption, we could conclude that the driving force behind the asymmetric information in Taiwan's cancer insurance market is adverse selection.

Furthermore, this paper investigates whether there is additional valuable information, and whether we could mitigate the severity of adverse selection by controlling this information. Our survey data provides us with a unique opportunity to collect such information and test it.[9] Our empirical results show that an individual's family cancer history constitutes valuable information in predicting the occurrence of cancer and the individual's behaviour in purchasing extended cancer insurance. Furthermore, it could help to mitigate the severity of the adverse selection problem.

An immediate implication of our finding is that requiring an individual's family cancer history may help insurers to price the insurance more accurately and to improve social welfare.[10] However, collecting this information may result in a loss of the individual's privacy. Thus, whether or not the insurance company should use such personal information is still debatable.

This paper consists of four parts. Following this introduction in the first part, the second part introduces the data and the methodology that we adopt. The third part presents the empirical results, and the fourth part concludes.

---

[6] The literature which uses the dynamic data to disentangle moral hazard and adverse selection includes: Abbring *et al.* (2003), Dionne *et al.* (2005), Dionne *et al.* (2007).

[7] Finkelstein and Poterba (2004) declares that moral hazard could play a smaller role in the annuity market, and they test for adverse selection using data for the annuity market in the U.K.

[8] This assumption is not so unreasonable. As described by Finkelstein and Poterba (2004), when they investigated the asymmetric information problem in the annuity market: "Moral hazard seems likely to play a smaller role in annuity markets, … While receipt of an annuity may lead some individuals to devote additional resources to life extension, we suspect that this is likely to be a quantitatively small effect". Although the person who has purchased extended cancer insurance could also pay less attention to taking care of his/her own health, and this neglectful behavior may cause him/her to contract cancer, it may also be a quantitatively small effect. Furthermore, this is the reason why we make such an assumption.

[9] The additional information that we collect in this paper is related to the health status and the family cancer history of an individual. These will be further introduced in the next section.

[10] Einav and Finkelstein (2010) inferred that the regulation that imposes restrictions that prohibit the insurance companies from making prices contingent on some risk factors could bring about a social welfare loss.

Data and methodology

Our questionnaires are collected from the policyholders of one big life insurance company in Taiwan.[11] The percentage of population purchasing extended cancer insurance in Taiwan is about 38 per cent nowadays. We distributed the questionnaires according to this percentage. We drew a sample of high coverage insured from this company's cancer insurance policy-holders, as well as a sample of low coverage insured from this company's life policyholders who had not purchased extended cancer insurance.[12] We also controlled the rate of contracting cancer in our sample according to the reality.[13]

We distributed 4,000 questionnaires, of which 3,788 were returned and 3,677 were valid. In our sample, the percentage of those contracting cancer was 1.28 per cent, and the percentage of interviewees with extended cancer insurance was 35.98 per cent. Both of them are close to the actual percentages in Taiwan.

The demographic information was also collected in our questionnaires, and included the interviewee's sexuality, age, marital status, educational back-ground, occupation, and residential area. In regard to the additional information, we collected two kinds of personal information. One was the personal health status, which was evaluated by the individual's frequency of NHI usage,[14] while the other had to do with the individual's family cancer history. The frequency of the individual's NHI usage is private information, which may not be revealed to any other people including the insurance

---

[11] The market share of this insurance company is the second largest in the life insurance market in Taiwan.

[12] In spite of this, it is still possible that in the group of life insurance policyholders that there are those who could, in fact, be covered by the other insurance company's cancer insurance contract. Through the questionnaire, we asked whether the interviewee had purchased extended cancer insurance in order to align this bias. After this modification, the percentage of those purchasing extended cancer insurance in our sample is about 36 per cent, which still approaches the percentage in reality.

[13] For this insurance company, the rate of contracting cancer for the policyholders who have the extended cancer insurance coverage is about 0.02, and the rate of contracting cancer for the policyholders who do not have the extended cancer insurance coverage is about 0.006. In our sample, the rate of contracting cancer in the sub-sample with the extended cancer insurance coverage is 0.018554 (with a standard deviation of 0.38515), and the rate of contracting cancer in the sub-sample without the extended cancer insurance coverage is 0.009566 (with a standard deviation of 0.27674).

[14] Although the best way to get the information about the health usage is to directly collect it from the provider, we cannot access such personal data regarding NHI usage from the provider. Hence, we can only hand collect each person's NHI usage by means of questionnaires.

companies. The individual's family cancer history also cannot be used for underwriting or pricing according to Taiwan's Insurance Act.

In terms of the additional information, only about 2.5 per cent of the interviewees have the worst health status, and 4.21 per cent of the interviewees have the worse health status.[15] Meanwhile, about 7.83 per cent of the interviewees are from families with a history of cancer.

Notice that identifying the timing of some of the variables is important to our research. For example, we detected the phenomenon of adverse selection (or advantageous selection) based on the conditional correlation between coverage and risk during the insurance contract. This makes the timing of contracting cancer and the timing of purchasing insurance important. Hence, we specifically asked about the timing in the questionnaire. Then, we discarded those observations where the insured purchased the insurance after contracting cancer.

In addition, when we employed the NHI usage frequency as the proxy for the individual's health status, the timing of the NHI usage was important for those who had contracted cancer. Thus, for the interviewees who had contracted cancer, we specifically asked about the NHI usage frequency before they contracted cancer. In this way we avoided the situation where the health status was affected by cancer and became an endogenous variable in our empirical model.

In Table 1, we compare the demographic characteristics of our survey data with Taiwan's real population structure. The structures of age, sex, education, marital status, and residential district basically approach Taiwan's real structures. However, we cannot compare such additional personal information because the real data are not available. Apart from this, the representation of our survey data appears convincing.

In Model 1, we test the asymmetric information problems through the conditional correlation between risk and coverage. The insurance company is not without any information regarding the insured. This is the reason why we test the asymmetric information problem through the conditional correlation instead of the simple correlation between risk and coverage. After controlling all the information that the insurance company holds for underwriting and pricing, the residual correlation between the risk and coverage can be inferred as evidence of asymmetric information.

---

[15] We define a dummy variable *hstatus_worst* for the persons with the worst health status, who used the NHI at least one time each week. We define a dummy variable *hstatus_worse* for the persons with worse health status, who used the NHI less than once in a week but at least once in a month. The worst and worse health statuses are compared with the better health status group who used the NHI less than once in a month.

Kili C. Wang *et al*
Asymmetric Information in Taiwan's Cancer Insurance Market
207

Table 1 Summary statistics of empirical sample and real population structure

Variable name	Empirical sample		Real population
	Mean	Std. dev.	
Explained variables			
cancer	0.0128	0.2012	0.0118
coverage_H	0.3598	0.4800	0.3800
Explanatory variables			
hstatus_worst	0.0250	0.1561	
hstatus_worse	0.0421	0.2009	
cancer_h	0.0783	0.2687	
sexf	0.5023	0.4996	0.4966
marri	0.5066	0.5001	0.4887
age_under25	0.3363	0.3640	0.3022
age_2535	0.1571	0.4725	0.1701
age_3545	0.2271	0.4413	0.1589
age_4565	0.2648	0.4191	0.2625
univ	0.3316	0.4991	0.3163
degree	0.0357	0.1857	0.0391
student	0.0994	0.2992	
teach	0.0445	0.2063	
mpf	0.0206	0.1419	
wl	0.2012	0.4010	
bl	0.0964	0.2952	
prof	0.0259	0.1590	
busi	0.0837	0.2770	
serv	0.0994	0.2992	
ins	0.1836	0.3872	
concern	0.0034	0.0584	
hous	0.0969	0.2959	
taip	0.1333	0.3950	0.1128
koa	0.0662	0.1869	0.0661
taichun	0.0630	0.3695	0.0464
tcm	0.1817	0.2740	0.1320
tpkl	0.1624	0.4400	0.1843
cct	0.1527	0.3598	0.1473
ycn	0.0927	0.2424	0.1027
iht	0.0418	0.1334	0.0448
kp	0.0989	0.1675	0.0919

Notes: The real population structure data is from Taiwan's Monthly Bulletin of Interior Statistics and the Directorate-General of Budget, Accounting and Statistics of Taiwan.

We do not have the real population structure comparison data regarding the hidden information such as health status (*hstatus_worst* and *hstatus_worse*) and family cancer history (*cancer_h*).

The categories of occupation are classified according to the requirements of our research. Hence, we do not have the real population structure comparison data regarding our occupation categories.

The first methodology we use is the pairwise Probit regression method adopted by Chiappori and Salanie (2000). The pairwise Probit regressions are as follows:

$$Prob\,(cancer_i = 1) = \Phi(X_i\beta_{cancer}). \tag{1}$$

$$Prob\,(coverage\_H_i = 1) = \Phi(X_i\beta_{cv}). \tag{2}$$

$cancer_i=1$ when the individual has been stricken by cancer, otherwise $cancer_i=0$; $coverage\_H_i=1$ when the individual has purchased extended cancer insurance, otherwise $coverage\_H_i=0$. β_{cancer} and β_{cv} are the vectors of the corresponding coefficients for each explanatory variable in each Probit regression. X_i is a vector of variables, which only includes all the variables that the insurance company uses for underwriting and pricing.[16] We can calculate their estimated residual as follows:

$$\hat{\varepsilon}_i^j = E\left(\frac{\varepsilon_i^j}{j_i}\right) = \frac{\phi(X_i\beta_j)}{\Phi(X_i\beta_j)}j_i$$

$$- (1-j_i)\frac{\phi(X_i\beta_j)}{\Phi(-X_i\beta_j)}, \quad j=cv, cancer \tag{3}$$

where Φ and ϕ represent the cumulative and density distribution functions of a standard normal distribution. On the basis of the two estimated residuals $\hat{\varepsilon}_i^{cancer}$ and $\hat{\varepsilon}_i^{cv}$, we can calculate a W statistic derived by Gourieroux $et\ al.$ (1987) as follows:

$$W_{jk} = \frac{\left(\sum_{i=1}^n \hat{\varepsilon}_i^j\hat{\varepsilon}_i^k\right)^2}{\sum_{i=1}^n (\hat{\varepsilon}_i^j)^2(\hat{\varepsilon}_i^k)^2}, \quad j, k = cancer, cv, j \neq k. \tag{4}$$

Under the hypothesis of $cov(\varepsilon_i^j, \varepsilon_i^k)=0$, the W_{jk} are distributed asymptotically as χ^2 (1). We reject the risk and coverage being conditionally independent and support the existence of asymmetric information while this statistic is significant. We can also calculate the conditional correlation coefficient of the above two estimated residuals and identify the driving force behind the asymmetric information problems. If the conditional correlation coefficient is significantly positive, under the assumption that moral hazard has little chance to play a role in the cancer insurance market, it is implied that the driving force is an adverse selection. On the contrary, the driving force is an advantageous selection.

---

[16] These variables are listed as the second part of the explanatory variables in Table 2.

Kili C. Wang *et al*
Asymmetric Information in Taiwan's Cancer Insurance Market
209

Table 2 Definitions of variables

Variable name	Definition
Explained variables	
cancer	A dummy variable that equals 1 when the individual has suffered from cancer after the individual has purchased the extended cancer insurance, otherwise 0.
coverage_H	A dummy variable that equals 1 when the individual has voluntarily purchased extended cancer insurance, otherwise 0.
Explanatory variables	
(1) First part	
hstatus_worst	A dummy variable that equals 1 when the individual used NHI at least one time each week, otherwise 0.
hstatus_worse	A dummy variable that equals 1 when the individual used NHI less than one time each week but at least one time each month, otherwise 0.
cancer_h	A dummy variable that equals 1 when the individual has a family cancer history, otherwise 0.
(2) Second part	
sexf	A dummy variable that equals 1 when the individual is female, otherwise 0.
marri	A dummy variable that equals 1 when the individual is currently married, otherwise 0.
age_under25	A dummy variable that equals 1 when the individual's age is under 25, otherwise 0.
age_2535	A dummy variable that equals 1 when the individual's age is equal to or above 25 and under 35, otherwise 0.
age_3545	A dummy variable that equals 1 when the individual's age is equal to or above 35 and under 45, otherwise 0.
age_4565	A dummy variable that equals 1 when the individual's age is equal to or above 45 and under 65, otherwise 0.
univ	A dummy variable that equals 1 when the individual has attended university or college, otherwise 0.
degree	A dummy variable that equals 1 when the individual has attended graduate school, otherwise 0.
student	A dummy variable that equals 1 when the individual is a student, otherwise 0.
teach	A dummy variable that equals 1 when the individual is a teacher, otherwise 0.
mpf	A dummy variable that equals 1 when the individual is in the military, the police force, or is a firefighter, otherwise 0.
wl	A dummy variable that equals 1 when the individual is a white-collar worker, otherwise 0.
bl	A dummy variable that equals 1 when the individual is a blue-collar worker, otherwise 0.
prof	A dummy variable that equals 1 when the individual's occupation is that of a technician or associate professional, otherwise 0.
busi	A dummy variable that equals 1 when the individual is a businessman, otherwise 0.
serv	A dummy variable that equals 1 when the individual's occupation is that of a service worker, or is engaged in shop and market sales, otherwise 0.
ins	A dummy variable that equals 1 when the individual works in the insurance industry, otherwise 0.

Table 2 (*continued*)

Variable name	Definition
concern	A dummy variable that equals 1 when the individual's occupation is related to health food sales, otherwise 0.
hous	A dummy variable that equals 1 when the individual's occupation is that of an unpaid family worker, otherwise 0.
taip	A dummy variable that equals 1 when the individual lives in Taipei city, otherwise 0.
koa	A dummy variable that equals 1 when the individual lives in Kaohsiung city, otherwise 0.
taichun	A dummy variable that equals 1 when the individual lives in Taichung city, otherwise 0.
tcm	A dummy variable that equals 1 when the individual lives in the Taoyuan, Hsinchu and Miaoli area, otherwise 0.
tpkl	A dummy variable that equals 1 when the individual lives in Taipei county or Keelung city, otherwise 0.
cct	A dummy variable that equals 1 when the individual lives in Taichung county, or the Changhua and Nantou area, otherwise 0.
ycn	A dummy variable that equals 1 when the individual lives in the Yunlin, Chiayi and Tainan area, otherwise 0.
iht	A dummy variable that equals 1 when the individual lives in the Ilan, Hualien and Taitung area, otherwise 0.
kp	A dummy variable that equals 1 when the individual lives in Kaohsiung county or the Pingtung area, otherwise 0.

The second methodology we use is similar to that of Dionne *et al.* (2001). In this model, we simultaneously estimate the probability of two endogenous variables, the occurrence of cancer ($cancer_i$) and the purchasing of extended cancer insurance ($coverage\_H_i$). They are conditional upon each other's endogenous variable, as well as other exogenous variables.

In the first stage, we run a probit regression for one of the endogenous variables ($cancer_i$) as follows:

$$Prob\,(cancer_i = 1|X_{1i}) = \Phi(X_{1i}\beta_1), \tag{5}$$

where β_1 is the parameter vector, and X_{1i} is an exogenous variable vector of each interviewee i. In this paper, the vector X_{1i} is exactly the same as vector X_i in Eqs (1) and (2).

In the second stage, we run the following probit regression for other endogenous variable ($coverage\_H_i$):

$$Prob(coverage\_H_i = 1|cancer_i, Pr\hat{o}b(cancer_i = 1|X_{1i}), X_{2i})$$
$$= \Phi(\beta_{cov,cancer}cancer_i + \beta_{estcancer}Pr\hat{o}b(cancer_i = 1|X_{1i}) + X_{2i}\beta_2), \tag{6}$$

where β_2 is the parameter vector of the exogenous vector X_{2i}. In this paper, X_{2i} is also consistent with X_i. $\beta_{estcancer}$ is the coefficient of estimated probability $Prôb(cancer_i)$ from the first stage. We test the conditional dependence between $cancer_i$ and $coverage\_H_i$ through the significance and the sign of the coefficient $\beta_{cov,cancer}$. If $\beta_{cov,cancer}$ is significantly positive, this means that there is significant positive conditional correlation between risk and coverage. Under our assumption, this also implies that the driving force behind the market's asymmetric information is adverse selection. On the contrary, it implies that the driving force is advantageous selection.

In Model 2, we add the additional information regarding the individual's health status to the explanatory variables. In Model 3, we add the additional information regarding the individual's family cancer history to the explanatory variables.[17] We use these two models to test whether the severity of asymmetric information could be mitigated after the hidden information is controlled.

We can identify the valuable additional information with the help of Models 2 and 3. We first examine whether these regressions imply that these variables bring valuable information regarding predicting the probabilities of cancer occurrence and purchasing cancer insurance. Moreover, we further observe the changes in the W statistic and the conditional correlation coefficient in our first methodology, as well as the change in $\beta_{cov,cancer}$ in our second methodology. If all the above terms are smaller or insignificant, they are evidence that the severity of the asymmetric information has been mitigated by controlling hidden information.

Empirical results

The first empirical result is concerned with the evidence for the asymmetric information problem. In Model 1 of Table 3, the W statistic in the pairwise Probit regressions supports the view that the conditional correlation between risk and coverage is significantly different from 0 at the 90 per cent confidence level. The conditional correlation coefficient is 0.0512, and is also significantly different from 0 at the 95 per cent confidence level. The empirical results under the two-stage method are consistent with those under the pairwise Probit regressions. In Model 1 of Table 4, the estimated coefficient $\beta_{cov,cancer}$ in the second-stage regression is 0.7016 and rejects the hypothesis of no conditional correlation at the 95 per cent confidence level. This is the evidence of asymmetric

---

[17] Vector X_i (as well as X_{1i} and X_{2i}) of Model 2 includes both the health status variables (*hstatus_worst* and *hstatus_worse*) and the second part of the explanatory variables in Table 2. Vector X_i (as well as X_{1i} and X_{2i}) in Model 3 includes both the family cancer history (*cancer_h*) and the second part of the explanatory variables in Table 2.

Table 3 The conditional correlation between purchasing extended cancer insurance and cancer occurrence in pairwise Probit regression methodology

	Model 1	Model 2	Model 3
W statistic	3.28967*	3.07913*	0.33364
	(0.0697)	(0.0793)	(0.2879)
Correlation coefficient of $\hat{\varepsilon}_i^{cv}$ and $\hat{\varepsilon}_i^{cancer}$	0.0512**	0.0494**	0.0092
	(0.0202)	(0.0249)	(0.5184)

Notes: There are no additional information variables included in the explanatory variables of Model 1. The explanatory variables of Model 2 also include the additional information variables of *hstatus_worst* and *hstatus_worse*. The explanatory variables of Model 3 also include the additional information variable of *cancer_h*.
**denotes 95 per cent confidence level, and *denotes 90 per cent confidence level. The *P*-values are in the parentheses.

Table 4 The coefficients of the Probit regressions in two-stage methodology

Explanatory variables	Model 1		Model 2		Model 3	
	First stage	Second stage	First stage	Second stage	First stage	Second stage
intercept	−4.4988	−4.3068	−4.6539	−4.3817	−4.7644	−4.3631
cancer		0.7016**		0.6824**		0.0321
Prob(cancer)		1.8630		1.5397		2.4772*
hstatus_worst			0.6162	0.1150		
hstatus_worse			0.5060	0.1662		
cancer_h					1.0642***	1.0155***
sexf	0.1042*	0.0180*	0.1172*	0.0160*	0.1391*	0.0213**
marri	−0.1156*	0.0364**	−0.0718*	0.0375*	−0.0394	0.0404
age_under25	−0.8962	−0.4405	−0.9688	−0.4466	−0.8872	−0.5060*
age_2535	−1.3122*	−0.1186	−1.3812**	−0.1246	−1.3294*	−0.1996
age_3545	−3.9639	0.0553	−4.0153	0.0498	−4.0833	−0.0361
age_4565	−0.8272	−0.0230	−0.8701	−0.0249	−0.8089	−0.0745
univ	−0.0669	−0.1004	−0.0519	−0.0967	−0.1322	−0.0981
degree	−2.8963	0.3581*	−2.8696	−0.3645**	−2.7887	0.3822**
student	−0.1710	0.1156	−0.1618	0.1182	0.0002	0.1078
teach	0.3276	0.0908	0.1403	0.0668	0.2358	0.0698
mpf	−2.6714	−0.1526	−2.6371	−0.1694	−2.5532	−0.1820
wl	−0.1077	0.1265	−0.1032	0.1249	0.0216	0.1209
bl	−2.5320	0.0066	−2.5319	0.0009	−2.5644	−0.0098
prof	0.6365	−0.0527	0.4545	−0.0721	0.7157	−0.0478
busi	0.3788	−0.1476	0.4031	−0.1472	0.7068	−0.1228
serv	0.4270	−0.0401	0.4008	−0.0399	0.5671	−0.0253
ins	0.3776	0.0288*	0.3363	0.0191*	0.3437	0.0262*
concern	1.5556*	−0.6428	1.5960*	−0.5894	1.6184*	−0.3185
hous	−3.0618	−0.0685	−3.3092	−0.0834	−2.9936	−0.1177

Table 4   (*continued*)

Explanatory variables	Model 1		Model 2		Model 3	
	First stage	Second stage	First stage	Second stage	First stage	Second stage
taip	2.9991	3.5563	3.1963	3.6320	3.0176	3.6922
koa	3.1220	3.2772	3.3057	3.3522	2.7631	3.4182
taichun	3.0484	4.3012	3.2254	4.3760	3.1753	4.4406
tcm	2.9120	4.0785	3.0308	4.1560	2.8080	4.2102
tpch	3.0297	4.3817	3.2151	4.4636	2.9419	4.5254
cct	2.9844	3.9953	3.1713	4.0689	2.9543	4.1282
ucn	0.0643	3.6514	0.2076	3.7147	−0.0151	3.7547
iht	0.2440	4.4261	0.4868	4.5019	−0.4156	4.5452
kp	−0.1343	3.7348	0.0419	3.8076	−0.1053	3.8544

Notes: There are no additional information variables included in the explanatory variables of Model 1. The explanatory variables of Model 2 also include the additional information variables of *hstatus_worst* and *hstatus_worse* in the second stage. The explanatory variables of Model 3 also include the additional information variable of *cancer_h* in the second stage.
***denotes 99 per cent confidence level; **denotes 95 per cent confidence level, and *denotes 90 per cent confidence level.

information and implies that the driving force is adverse selection, if moral hazard plays only a small role in the market.

The additional information regarding the individual's health status is then added in Model 2. Under the pairwise Probit method in Table 5, both estimated coefficients of *hstatus_worst* and *hstatus_worse* are insignificantly different from 0. By observing their marginal effect, we calculate the change in the probability of contracting cancer, as well as in that of purchasing extended cancer insurance caused by those two variables. We find that those who have the worst health status (have the worse health status) have a 0.094 per cent (0.001 per cent) higher probability of contracting cancer than others. Both differences are not significantly different from 0. On the other hand, those persons who are with the worst health status (with the worse health status) have a 0.058 per cent (0.007 per cent) higher probability of purchasing extended cancer insurance than others. Both differences are also not significantly different from 0.

These results could indicate that the health status revealed by the frequency of NHI usage may imply the chronic disease condition of one person, and the person who contracts some chronic disease may not contract the cancer. Hence, the person who is known to have the worst or worse health status due to some chronic disease may also not necessarily be more intent on purchasing cancer insurance than others. Accordingly, after we include the variables *hstatus_worst* and *hstatus_worse*, the adverse

Table 5 The coefficients of the Probit regressions for purchasing extended cancer insurance and cancer occurrence in pairwise Probit regression methodology

Explanatory variables	Model 1		Model 2		Model 3	
	Probit regression (1)	Probit regression (2)	Probit regression (1)	Probit regression (2)	Probit regression (1)	Probit regression (2)
intercept	−4.4988	−4.2489	−4.6539	−4.3546	−4.7644	−4.3565
hstatus_worst			0.0616	0.0148		
hstatus_worse			0.0506	0.0198		
cancer_h					1.0642***	1.3301***
sexf	0.1042*	0.0250*	0.1172*	0.0217*	0.1391*	0.0215**
marri	−0.1156*	0.0412*	−0.1018*	0.0415*	−0.0394	0.0414*
age_under25	−0.8962	−0.5270**	−0.9688	−0.5185**	−0.8872	−0.5195**
age_2535	−1.3122**	−0.2250	−1.3812**	−0.2142	−1.3294*	−0.2135
age_3545	−3.9639	−0.0654	−4.0153	−0.0519	−4.0833	−0.0513
age_4565	−0.8272	−0.0940	−0.8701	−0.0846	−0.8089	−0.0845
univ	−0.0669	−0.1014	−0.0519	−0.0972	−0.1322	−0.0978
degree	−2.8963	0.3793**	−2.8696	0.3841**	−2.7887	0.3848**
student	−0.1710	0.1006	−0.1618	0.1044	0.0002	0.1071
teach	0.3276	0.1017	0.1403	0.0688	0.2358	0.0696
mpf	−2.6714	−0.1709	−2.6371	−0.1888	−2.5532	−0.1873
wl	−0.1077	0.1169	−0.1032	0.1162	0.0216	0.1169
bl	−2.5320	−0.0099	−2.5319	−0.0145	−2.5644	−0.0147
prof	0.6365	−0.0152	0.4545	−0.0485	0.7157	−0.0457
busi	0.3788	−0.1207	0.4031	−0.1240	0.7068	−0.1225
serv	0.4270	−0.0218	0.4008	−0.0260	0.5671	−0.0253
ins	0.3776	0.0367*	0.3363	0.0238*	0.3437	0.0227*
concern	1.5556*	−0.2966	1.5960*	−0.2894	1.6184*	−0.2897
hous	−3.0618	−0.1166	−3.3092	−0.1273	−2.9936	−0.1256
taip	2.9991	3.6152	3.1963	3.7033	3.0176	3.7019
koa	3.1220	3.3546	3.3057	3.4411	2.7631	3.4362
taichun	3.0484	4.3617	3.2254	4.4484	3.1753	4.4491
tcm	2.9120	4.1299	3.0308	4.2199	2.8080	4.2202
tpch	3.0297	4.4422	3.2151	4.5367	2.9419	4.5352
cct	2.9844	4.0520	3.1713	4.1384	2.9543	4.1369
ucn	0.0643	3.6858	0.2076	3.7622	−0.0151	3.7616
iht	0.2440	4.4632	0.4868	4.5548	−0.4156	4.5507
kp	−0.1343	3.7772	0.0419	3.8638	−0.1053	3.8615

Notes: Probit regression (1) is regressed on *cancer*. Probit regression (2) is regressed on *coverage*. There is no additional information variable included in the explanatory variables of Model 1. The explanatory variables of Model 2 also include the additional information variables of *hstatus_worst* and *hstatus_worse*. The explanatory variables of Model 3 also include the additional information variable of *cancer_h*.

***denotes 99 per cent confidence level; **denotes 95 per cent confidence level, and *denotes 90 per cent confidence level.

selection is hardly mitigated. In Tables 1 and 3, regardless of which methodology is used, we can see that the conditional correlation between coverage and risk in Model 2 is not significantly different from that in Model 1. In other words, there is no evidence to support the view that the individual's health status as we defined it contains valuable hidden information in the cancer insurance market.

The additional information we employed in Model 3 is the individual's family cancer history. While we predict the probabilities of cancer occurring and purchasing cancer insurance in Table 5, the estimated coefficients of the family cancer history variable are significantly positive at the 99 per cent confidence level in the regressions. We calculate the marginal effect of this additional variable. Those individuals who come from families with a history of cancer are found to have a 0.35 per cent higher probability of contracting cancer, with this inference being significant at the 99 per cent confidence level. Likewise, those individuals with a family cancer history have a 0.042 per cent higher probability of being willing to purchase extended cancer insurance, and this inference also reaches the 99 per cent confidence level. This preliminary observation indicates that the individuals' family cancer history could contain valuable information in improving the problem of adverse selection.

While we control for family cancer history in Model 3 under the pairwise Probit regression method, the value of the W statistic clearly declines from 3.28967 in Model 1 to 0.33364. The W statistic here cannot reject the hypothesis that the conditional correlation is different from 0 under any statistically significant level. The conditional correlation coefficient also dramatically decreases from 0.0512 in Model 1 to 0.0092. The conditional correlation coefficient here also cannot reject the hypothesis that there is no conditional correlation. Under the two-stage method, in Model 3, the value of the conditional correlation estimated coefficient $\beta_{cov,cancer}$ is 0.0321, which is much smaller than that in Model 1 and it also cannot reject the null hypothesis of no correlation ($H_0:\beta_{cov,cancer}=0$). Both of these two methodologies result in a consistent outcome in that the severity of adverse selection is dramatically mitigated after the family cancer history is controlled for.

When we compare the outcomes of Models 3 and 1, we cannot reject the conditional correlation in Model 3, but we can reject the conditional correlation in Model 1. The difference between these two models is that we add additional information regarding the individual's family cancer history into the explanatory variables of Model 3. This additional information absorbs the correlation between risk and insurance in this insurance market.

This evidence supports the view that the family cancer history variable provides valuable information in predicting cancer risk and the willingness to purchase extended cancer insurance. Furthermore, it can also help to reduce

the severity of adverse selection by controlling the valuable information. These conclusions remain robust under different tests of methodologies.

Conclusion

This paper investigates the asymmetric information problems in Taiwan's cancer insurance market by using survey data. When we treat the individuals who also purchase extended cancer insurance other than the coverage provided by the NHI as the insured with high insurance coverage, we find evidence of significant positive conditional correlation between risk and coverage. While the possibility of moral hazard playing a role in this market is only minor, we can accordingly infer that this is evidence of adverse selection. Hence, this is the first study that finds empirical support for the existence of asymmetric information in the cancer insurance market, and indicates that the driving force could be adverse selection.

Of course, if moral hazard plays a significant role in the cancer insurance market, we can only conclude that asymmetric information exists, but cannot identify its source. This possibility can only be tested by having further information regarding a person's behaviour or efforts to avoid contracting cancer, such as not smoking, exercise, and eating healthy food, etc., both before and after purchasing cancer insurance.

One of the advantages of our paper in testing the asymmetric information problems through survey data is that we can hold the real accident occurrence information instead of using claims information as the proxy variable.[18] In this way, we could avoid a false positive conditional correlation between risk and coverage. Hence, our adverse selection inference is convincing from this point of view. The other advantage of our survey data is that we can not only have data for the individual who has purchased insurance, but we can also have data for the individual who has not purchased insurance. This enables us to investigate the asymmetric information problems among the individuals who are both covered and not covered by extended cancer insurance.

With the help of hand-collected data, we have a unique opportunity to approach some personal information of the interviewee. We collect two kinds of additional information, namely, the individual's health status and family

---

[18] Empirical studies often use data obtained from the insurance company. They tend to only have data on the claim variable instead of the real accident occurrence variable. As was pointed out by Chiappori and Salanie (2000), a high coverage contract could tend to encourage claims for small accidents in contrast to a low coverage contract. Hence, using the claim variable as a proxy for the accident variable will tend to give rise to a spurious positive correlation between risk and coverage.

cancer history, neither of which the insurance companies are permitted to use by law to underwrite and price their cancer insurance.

We find that the additional information regarding the individual's health status provides little help in predicting the probability of contracting cancer or the willingness to purchase extended cancer insurance. It also does not offer much help in mitigating the problem of adverse selection in the insurance market. However, we find that the additional information regarding the individual's family cancer history is valuable. The individual who has a family with a history of cancer has a higher probability of contracting cancer, and is also more willing to purchase cancer insurance. Furthermore, we find from our empirical results that controlling this additional information could help to reduce the severity of adverse selection.

Our findings seem to imply that the existence of adverse selection in Taiwan's cancer insurance market arises due to the government's failure. That the regulation prevents the insurance companies from using such valuable information as a family's cancer history is worth further discussion. A change in the regulations to reveal the "family cancer history" could lead to a gain in social welfare. Some recent studies including Einav and Finkelstein (2010) have highlighted this problem. The amount of welfare loss from adverse selection caused by this regulatory restriction may even be measured with the help of some additional information such as the proportion of uninsured people and the demand elasticity of insurance. Nevertheless, this interesting issue could be left to future research provided that the above information is available.

Of course, while adverse selection could be mitigated and social welfare improved by using such additional information, the trade-off is that its use sacrifices the privacy of an individual. This raises an important question as to whether some of the valuable personal information should be held by the insurance company and used for underwriting or pricing.[19] This is one issue needed in relation to asymmetric information, and is a topic that deserves careful investigation in the future.

Acknowledgements

---

[19] Chiappori (2006) and Hoy and Witt (2007) have discussed how the social welfare could be improved if the genetic test information were allowed to be used. Although the additional personal information in this paper, such as the family cancer history of the insured, is not exactly the same as the information from a genetic test, the dilemma that results from trying to resolve the asymmetric information problem and maintaining privacy is similar.

References

Abbring, J.H., Chiappori, P.A., Heckman, J.J. and Pinquet, J. (2003) 'Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish?' *Journal of the European Economic Association* **1**(2–3): 512–521.

Cardon, J.H. and Hendel, I. (2001) 'Asymmetric information in health insurance: Evidence from the national medical expenditure survey', *Rand Journal of Economics* **32**(3): 408–427.

Cawley, J. and Philipson, T.J. (1999) 'An empirical examination of information barriers to trade in insurance', *American Economic Review* **89**(4): 827–846.

Chiappori, P.A. (2006) 'The welfare effects of predictive medicine', in P.A. Chiappori and C. Gollier (ed.) *Competitive Failures in Insurance Markets: Theory and Policy Implications*, Cambridge, MA: MIT Press.

Chiappori, P.A. and Salanie, B. (2000) 'Testing for asymmetric information in insurance markets', *Journal of Political Economy* **108**(1): 56–78.

Cohen, A. (2005) 'Asymmetric information and learning: Evidence from the automobile insurance market', *The Review of Economics and Statistics* **87**(2): 197–207.

Cohen, A. and Einav, L. (2007) 'Estimating risk preference from deductible choice', *American Economic Review* **97**(3): 745–788.

de Meza, D. and Webb, D.C. (2001) 'Advantageous selection in insurance markets', *Rand Journal of Economics* **32**(2): 249–262.

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): 444–455.

Dionne, G., Maurice, M., Pinquet, J. and Vanasse, C. (2005) *The role of memory in long-term contracting with moral hazard: Evidence in automobile insurance*, working paper, SSRN #764705.

Dionne, G., Michaud, P. and Dahchour, M. (2007) *Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from France*, working paper.

Einav, L. and Finkelstein, A. (2010) *A graphical analysis of selection in insurance markets Working paper*.

Finkelstein, A. and McGarry, K. (2006) 'Multiple dimensions of private information: Evidence from the long-term care insurance market', *American Economic Review* **96**(4): 938–958.

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): 183–208.

Gourieroux, C., Monfort, A., Renault, E. and Trognon, A. (1987) 'Generalized residuals', *Journal of Econometrics* **34**(1): 5–32.

Hoy, M. and Witt, J. (2007) 'Welfare effects of banning genetic information in the life insurance market: The case of BRCA1/2 genes', *Journal of Risk and Insurance* **74**(3): 523–546.

Huang, R.J., Liu, Y. and Tzeng, L.Y. (2006) *Hidden overconfidence and advantageous selection*, working paper.

Huang, R.J., Muermann, A. and Tzeng, L.Y. (2008) *Hidden regret in insurance markets: Adverse and advantageous selection*, working paper.

Koufopoulos, K. (2003) *Asymmetric information, heterogeneity in risk perceptions and insurance: An explanation to a puzzle*, working paper.

Puelz, R. and Snow, A. (1994) 'Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market', *Journal of Political Economy* **102**(2): 236–257.

Richaudeau, D. (1999) 'Automobile insurance contracts and risk of accident: An empirical test using French individual data', *The GENEVA Papers on Risk and Insurance Theory* **24**(1): 97–114.

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**(4): 629–649.

Saito, K. (2006) 'Testing for asymmetric information in the automobile insurance market under rate regulation', *Journal of Risk and Insurance* **73**(2): 335–356.

Shavell, S. (1979) 'On moral hazard and insurance', *Quarterly Journal of Economics* **93**(4): 541–562.
Sonnenholzner, M. and Wambach, A. (2006) *On the role of patience in an insurance market with asymmetric information*, *working paper*.

About the Authors

Kili C. Wang is Associate Professor at the Department of Insurance, Tamkang University.

Jin-Lung Peng is Associate Professor at the Department of Risk Management and Insurance, Shih Chieh University.

Yi-Yun Sun is a Master's degree holder from Graduate Institute of Finance and Insurance, Shih Chien University.

Yao-Chia Chang holds a Master's degree from the Graduate Institute of Finance and Insurance, Tamkang University.