International Journal of Information and Management Sciences

http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information and Management Sciences

32 (2021), 195-208. DOI:10.6186/IJIMS.202109_32(3).0001

Who Purchases Private Health Insurance under Taiwan's National Health Insurance Program: An Empirical Study from Latent Class Analysis

Joseph J. Tien[1] *and Wei-Hua Tian*[2]
[1]Tamkang University and [2]National Cheng Kung University

Keywords

Latent class analysis
Risk preference
Private health insurance
Inpatient service utilization

Abstract.

The purpose of this study is to examine who purchases private insurance in Taiwan based on latent class analysis (LCA) framework. The 2005 Nation Health Interview Survey (NHIS) linked to the 2006 and 2007 National Health Insurance Research Claim Database (NHIRD) have been used to identify individual's risk preference classification by LCA. We further utilize a bivariate probit regression to investigate the relationship between each latent risk preference classification and purchase of private health insurance, and the subsequent inpatient service utilization. Empirical results indicate that compared with the high cautiousness class, individuals in the low cautiousness with smoking behavior class tend to less likely to purchase private health insurance; however, individuals in the low cautiousness with smoking behavior class and private health insurance tend to more likely to utilize inpatient services. Our empirical results provide information about how insurance companies can perform a target market strategy on their potential customers.

1. Introduction

The risk attitude plays a prominent role in decision-making under the uncertain environment. Rothschild and Stiglitz [20] provided the theoretical model to demonstrate that individual with higher expected claims intend to purchase more insurance than those with lower expected claims. That is to say, higher-risk type individuals tend to buy more insurance. Asymmetric information (whether adverse selection or ex ante moral hazard) which documents the relationship between the individual's risk-type and insurance purchase has been regarded as important issues after Rothschild and Stiglitz [20]. To follow the theoretical model, numerous empirical studies had attempted to test whether or not asymmetric information between insurance coverage and risk occurrence exists under different types of insurance markets such as health, automobile, long-term care, and pension insurance markets. However, the empirical results are mixed. Some studies supported the evidence that asymmetric information exists (see Puelz and Snow [19] Cohen [8]) while others did not support it (see Cawley and Philipson [6] Chiappori et al. [7]).

In statistical terms, asymmetric information can be tested by the positive correlation between the residuals of a two-equation model of insurance demand and risk of occurrence (see Chiappori et al. [7]). Since the empirical results are mixed, there are studies had paid more attention to discuss the limitations of the positive correlation test. For example: Finkelstein and McGrary [13] had pointed out that when individuals have private information such as risk preferences other than risk types that may affect the demand for insurance. The validity of the positive correlation test may be problematic. As a result, there were studies had examined the relationship between individuals' insurance preference, risk type, and insurance demand under different type of insurance market (see Finkelstein and Poterba [14] Finkelstein and McGrary [13] Cutler et al. [9] Li Donni [18] Sloan et al. [22] Alessie et al. [2]). Their results had confirmed private information such as risk preferences may affect the demand for insurance. Specifically, Finkelstein and McGrary [13] utilized Health and Retirement Study (HRS) in the United States to empirically examine the implications of multiple dimensions of private information in the long-term care insurance markets. Their results demonstrate that the standard test of asymmetric information is invalid which is to say the positive correlation between insurance coverage and risk occurrence was not found. However, the insurance market may still suffer from asymmetric information. Cutler et al. [9] used the individual-level data from the HRS to examine the relationship between risky behaviors, insurance purchases, and risk occurrences in life insurance, acute health insurance, annuities, long-term care insurance, and Medicare Health insurance markets. Their results indicate in all five markets, individuals who engage in risky behaviors such as smoking or drinking or who do not take measures to reduce risk such as utilizing preventive health activities or wearing a seat belt while driving are systematically less likely to hold any one of these insurances and are having higher expected claims for life insurance and long-term care insurance and lower expected claims for annuities, Medigap and acute health insurance. Li Donni [18] utilized HRS and latent class analysis (LCA) to examine the relationship between unobserved risk preferences and insurance purchase decisions. Results confirm that heterogeneity in risk preferences affect multiple insurance demands.

As many previous studies had investigated the relationship between individual's private information and the demand for insurance and confirmed that the preference heterogeneity may affect the insurance demand (see Finkelstein and Poterba [14] Finkelstein and McGrary [13] Cutler et al. [9] Li Donni [18] Sloan et al. [22] Alessie et al. [2]), our paper addresses this issue by utilize a LCA framework to account for unobserved heterogeneity across individuals, producing an effective model of observed risk preference classifications. LCA is widely used in studies regarding medical utilization and health care field (see Deb and Trivedi [10] Bago dUva [4] Bago dUva [5] Greene et al. [15] Scharoun- Lee et al. [21] Lafortune [16] Laska et al. [17]). However, few papers applied LCA to the insurance field (see Li Donni [18]). LCA is usually used to find subgroups or subtypes of cases in multivariate categorical data. A latent class model translates a set of observed multivariate variables into a set of latent variables. A latent class group is characterized by a pattern of conditional probabilities that indicate the probabilities that latent variables take on certain values. Thus, using the individual-level data from the 2005 National Health Interview Survey (NHIS) and the claim data of the inpatient service utilization from the 2006 and 2007 National Health Insurance Research Database (NHIRD) in Taiwan, we utilize characteristics of individual's risk behaviors such as smoking and drinking behaviors as risky-factor proxies to identify the entire samples into different latent classes of risk preference. Then, we utilize a bivariate probit regression to investigate the relationship between each latent class of risk preference and purchase of private health insurance (PHI) and the subsequent inpatient service utilization. We also perform the positive correlation test to check asymmetric information. Taiwan's National Health Insurance (NHI) program was implemented in March 1995 with the goal to provide every citizen with equal access to comprehensive medical services regardless of their socio-economic status. Our analysis thus provides an evidence for the supplementary PHI market in Taiwan. Based on our results, insurance companies can perform a target market strategy on their potential customers who may purchase PHI in the future.

The remainder of this study is organized as follows. Section 2 presents the conceptual framework. Section 3 presents the data resource and the empirical model. Section 4 provides the results. Section 5 makes the conclusions and discussions.

2. Conceptual Framework

Previous literature often used individual-level data to conduct the so-called "positive correlation test" empirically (see Puelz and Snow [19] Cawley and Philipson [6] Chiappori et al. [7] Cohen [8]). While the positive correlation test often assumes a positive correlation prediction of asymmetric information models between insurance coverage and risk occurrence; however, some findings do not support the positive correlation (see Cawley and Philipson [6] Chiappori et al. [7]). Indeed, the literature always emphasizes the role of private information about individual's risk in determining the decision of purchasing insurance. Therefore, following the idea of the existence of unobserved heterogeneity in risk preferences that may affect the equilibrium of insurance market with asymmetric information. There are studies had paid more attention to discuss the relationship between individuals' unobserved heterogeneity in risk preferences and the insurance demand (see Finkelstein and Poterba [14] Finkelstein and McGrary [13] Cutler et al. [9]). For example: Finkelstein and McGarry [13] pointed out that when individuals have private information such as risk preference, it may possibly affect insurance demand. Cutler et al. [9] and Einav et al. [11] also suggest that when conducting the positive correlation test, one needs to consider the risk preference heterogeneity. Finkelstein and Poterba [14] used U.K. annuity data to provide more evidence of the existence of unobservable characteristics among individuals in the decision to purchase annuity. Their findings suggested the possible existence of risk preference-based selection effect in insurance market.

Thus, in this paper, we firstly use LCA to identify latent classes of risk preference i.e. individuals' unobserved heterogeneity in risk preferences. We then investigate the relationship between each latent classes of risk preference and demand for health insurance. Specifically, studies indicate that despite the risk type, if unobserved preference is positively with the insurance demand and negatively with the risk occurrence, the correlation between insurance coverage and risk occurrence can be negative. This is the so-called "advantageous" selection. Thus, unobserved heterogeneity preference may offset the adverse selection and obscure the expected positive correlation between insurance coverage and risk occurrence (see Finkelstein and McGrary [13] Cutler et al. [9]). We can address this by investigate the relationship between each latent class of risk preference and the subsequent inpatient utilization and perform the positive correlation test.

3. Empirical Model

3.1. Data resource and study sample

Our data were obtained from the following two secondary data resources. The first one is the 2005 NHIS. This database adopted a face-to-face interviews and a multi-stage stratified systematic sampling design method. The NHIS was sourced from the National Health Research Institutes (NHRI) and the Health Promotion Administration (HPA) within the Ministry of Health and Welfare in Taiwan. The NHIS provides nationwide detailed population information on a series of individual characteristics, including age, gender, marital status, educational attainment, income, and detailed information on personal health conditions and diseases status.

The second database used in this study was released by the NHRI directly. The NHRI created this database by drawing data from the 2006 and 2007 NHIRD. The NHIRD database is one of the largest and most comprehensive population-based claim data sources that include registries of medical facilities contracted with the National Health Insurance Administration, board-certified physicians, a monthly claim summary for all inpatient/outpatient claims, details of all inpatient/outpatient orders, and expenditure on prescriptions dispensed at contracted pharmacies. Therefore, by linking these two databases, we can examine the relationship between each latent risk class, purchase of PHI and the subsequent inpatient service utilization.

3.2. Empirical model and variables of interest

3.2.1. Latent class of risk preference

Finkelstein and McGrary [13] demonstrated the existence of various dimensions of private information in the long-term care insurance market. They used the utilization of preventive healthcare activities or wearing a seat belt while driving as proxies for the risk attitudes and found that the individuals who behave more cautiously are more likely to own insurance and less likely to use long-term care services. Cutler et al. [9] used five measures of behaviors that capture individuals risk preferences including drinking, smoking, job-based mortality risk, preventive health care utilization, and use of seat belts. Therefore, to construct the risk preference indicators, it is necessary to use observable patterns of risk behaviors followed by previous literature. We thus choose six indicators from the survey questionnaire to construct latent risk classes including preventive health service utilization, verifying prescription drug name from the prescription bag, smoking, drinking, wearing a seat belt while driving, and wearing a helmet while riding a bike. Specifically, preventive health service utilization is coded as 1 if the individual had utilized the free adult preventive care services under the NHI program last year or within the previous 3 years (The services are provided once a year for citizens aged 65 years or older and once every three years for citizens between age 40-64 years). Verifying prescription drug name from the prescription bag is coded as 1 if the individual checks prescription drug name from the prescription bag when receiving it almost every time. Smoking is coded as 0 if the individual smokes every day. Drinking is coded as 0 if the individual drinks every day and is drunk most of the time. Wearing a seat belt or helmet is coded as 1 if the individual does not wear it every time when individual drives or rides a motorcycle.

Subsequently, we use a series of model fit criterion include Pearson chi-square, likelihood ratio chi-square, Akaike information criterion (AIC), Bayesian/Adjusted Bayesian information criterion (BIC/Adj BIC) to identify the latent risk classes. Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (LMR) was used to compare the model fit between sequential classes. Table 1 presents the model fit indexes, based on these six indicators, we were able to identify three classes in the following analysis.

Table 1: Latent class model fit indexes.

	2 classes	3 classes	4 classes
Pearson Chi-square	202.924	70.090	51.79
LR Chi-square	195.140	70.644	50.391
Chi-square df	50	43	36
Log likelihood	-12676.485	-12614.237	-12604.110
AIC	25378.970	25268.475	25262.221
BIC	25463.487	25398.500	25437.755
Adj-BIC	25422.177	25334.947	25351.959
Entropy	0.636	0.795	0.586
Vuong-Lo-Mendell-Rubin Likelihood Ratio Test	1 Versus 2 Classes	2 Versus 3 Classes	3 Versus 4 Classes
LMR probability	0.0000	0.0000	0.0200

Note1: the smaller the AIC or BIC value is, the better the model fit is.
Note2: LMR test compares the improvement in model fit between sequential classes, $p < 0.0001$ is considered to have a good model fit.

Table 2 further presents the statistics among the three latent classes of risk preference and each indicator. The first column is the sample proportion based on the six indicators. There are about 38.10 percent of individuals who report that they had preventive health service utilization last year or within the previous 3 years, while 44.67 percent of individuals had verified prescription drug name from the prescription bag.

Individuals who report that they do not drink to drunk every day or smoke are about 75 percent and 97.97 percent, respectively. There are about 92.75 percent and 89.11 percent of individuals who report that they wear a seat belt or a helmet every time when the individual take a car or ride a motorcycle. The second, third and fourth columns represent the conditional probabilities of each indicator in each class as if the individual's response is yes. Specifically, the first class is considered to be the "high cautiousness" class. The conditional probabilities of each indicator in this class are generally high relative to other classes. For example: there are about 41.2 percent and 48.1 percent that individuals report that they had preventive health service utilization last year or within the previous 3 years and verified prescription drug name from the prescription bag, respectively. Class 2 is considered to be the "low cautiousness without wearing a seat belt or helmet" class. In this class, the conditional probabilities of the indicators relating to the wearing a seat belt and helmet are the lowest among each class. There are about 33.4 percent and 29 percent of individuals who report that they do not wear a seat belt or a helmet every time when taking a car or riding a motorcycle, respectively. Class 3 is considered to be the" low cautiousness with smoking behavior" class. There are about 100% of individuals who report that they smoke every day in this class (smoking is coded as 0 if the individual smokes every day).

Table 2: Latent classes of risk preference.

	Sample Proportion	Class1	Class2	Class3
Preventive Care Utilization	38.10%	41.2%	31.6%	12.7%
Verifying	44.67%	48.1%	21.9%	26.7%
Drink	75.00%	99.5%	91.3%	87.2%
Smoke	97.97%	83.7%	61.8%	0
Seat Belt	92.75%	96.9%	33.4%	91.8%
Helmet	89.11%	93.1%	29.0%	90.0%

Note: Class 1:"high cautiousness"; Class 2: "low cautiousness without wearing a seat belt or helmet" and Class 3: "low cautiousness with smoking behavior".

3.2.2. Regression model

We utilize a bivariate probit regression to investigate the relationship between each latent class of risk preference and purchase of PHI and the subsequent inpatient service utilization. A bivariate probit model can simultaneously estimate the demand for health insurance and the utilization of inpatient services while avoiding possible endogeneity. Further, by utilizing the bivariate probit model, we can conduct the "positive correlation" test to confirm that if asymmetric information presents in Taiwan's PHI market.

The following equations presents the bivariate probit model. Equation (3.1) describes the demand for PHI. The explanatory variables that have been widely adopted in previous studies is utilized in this equation (see Tian et al. [23]). Equation (3.2) models the utilization of inpatient services, featuring the explanatory variables used in Andersen's health service utilization behavior model (see Andersen and Newman [3] Aday and

Andersen [1]). The dependent variables for demand for PHI and inpatient care service utilization equations are dichotomous. The equations are as follows:

$$\text{PHI} = \beta_1 \text{CLASS}_i + \beta_2 X_j + \varepsilon_1 \tag{3.1}$$
$$\text{Inpatient} = \gamma_1 \text{CLASS}_i + \gamma_2 \text{CLASS}_j * \text{PHI} + \gamma_3 \text{PHI} + \gamma_4 X_k + \varepsilon_2 \tag{3.2}$$
$$[\varepsilon_1, \varepsilon_2] \sim \text{BVN}[(0,0), \sigma_1^2, \sigma_2^2, \rho]$$

where PHI is a dichotomous variable that indicates whether the individual had purchased PHI (PHI $=1$ indicates that the individual had purchased PHI; otherwise it is 0); Inpatient is a dichotomous variable that indicates whether the individual had received inpatient care services in 2006 and 2007. X_j and X_k are vectors of the explanatory variables. In equation (3.1): the key independent variable is CLASS. There are 3 latent classes of risk preference in the model: the "high cautiousness"; "low cautiousness without wearing a seat belt or helmet" and" low cautiousness with smoking behavior" classes. X_j includes personal characteristics, such as age, gender, marital status, educational attainment, monthly household income, residential location, ethnicity, health status and the presence of chronic diseases, such as hypertension, diabetes, asthma, stroke, kidney disease, heart disease, gout, and stomach disease etc.

We further consider that the effect of PHI on the demand for health services as well as holding PHI is endogenous. We thus propose an instrumental variable (IV) approach. A valid instrument must be highly correlated with the purchase of PHI, but unrelated to the utilization of inpatient service. Therefore, we employ two instruments: looking for food and nutrition labeling and asking prescription drug information from the pharmacist to control for endogeneity.

In equation (3.2): the key independent variable is CLASS*PHI, where CLASS*PHI is the interaction of each latent class and PHI status. X_k includes personal characteristics and the presence of chronic diseases etc. ε_1 and ε_2 are error terms that are assumed to be correlated across the two equations when considering endogeneity with bivariate normal distributions. σ_1 and σ_2 are standard deviations. Specifically, ρ is the correlation coefficient of ε_1 and ε_2 and its significance is tested by means of the likelihood-ratio test. We can thus perform the positive correlation test to check asymmetric information.

4. Empirical Results

Table 3 presents the first equation estimation results of the bivariate probit model: the health insurance demand. Since self-rated health status is subjective (see Wu [24]) and may affect the demand for health insurance and the utilization of inpatient service, in order to check the robustness of our estimates, we have constructed three models as below. Model 1 uses self-reported health status of the individuals as controlling variables, model 2 uses disease status of the individuals as controlling variables, and model 3 uses both sets of variables as controlling variables.

In general, the empirical results of the three models all show that relative to the high cautiousness class, individuals in the low cautiousness with smoking behavior class tend to have lower probability to buy PHI. Furthermore, in model 1, 2, and 3, other explanatory variables that positively significantly affect the likelihood of health insurance demand, such as female, younger age, higher monthly income level, higher educational attainment, and individuals located in the southern area compared to the northern area of Taiwan. However, being mainlander relative to other ethnicity such as aboriginals and individuals who looked for food and nutrition labeling were found to negatively significantly affect the likelihood of purchasing PHI. Specifically, in model 1 and 3, as regard to health status, apart from those reporting a very good health condition, there is a positive likelihood of purchasing PHI for those individuals reporting that their health condition is fair. In model 2 and 3, individuals with chronic diseases such as hypertension and diabetes were found to negatively significantly affect the likelihood of purchasing PHI.

Table 4 presents the second equation estimation results of the bivariate probit model: the inpatient services utilization. The empirical results of the three models all demonstrate that relative to the high cautiousness class, individuals in the low cautiousness with smoking behavior class tend to have higher probability to utilize inpatient service in the subsequent 2 years once they are insured. Furthermore, in Model 1,2 and 3, other explanatory variables that positively significantly affect the likelihood of inpatient service utilization include individuals with PHI and male. Variables that negatively significantly affect the likelihood of inpatient service utilization include younger age, being single, higher monthly income level and educational attainment. Specifically, in Table 1, 2, and 3 health status is very bad compared to that of very good, individuals with chronic diseases such as hypertension, diabetes, asthma, heart disease, gout, stomach disease, osteoporosis, liver disease, cancer, mental disease were found to positively significantly affect the likelihood of inpatient service utilization.

Finally, we find that the correlation coefficients ρ that presented in the three models are all negatively significant. The likelihood-ratio test statistics of chi-square are 2.883 (p-value = 0.09), 2.753 (p-value = 0.097), and 3.293 (p-value = 0.07), respectively for model 1, 2 and 3.

5. Conclusion and Discussion

Apart from previous studies which only used separated risk factors in analyzing the demand for insurance, our paper contributes to the literature in two ways. First, this study adopts LCA to identify risk behaviors into three major latent classes of risk preference: the "high cautiousness"; "low cautiousness without wearing a seat belt or helmet", and "low cautiousness with smoking behavior". Individuals in each class show different risk preferences based on their risk-relating behaviors. LCA translates the six observed risk variables into a set of latent variables and identifies the mutually exclusive classes of risk preference where within-group risk preference differences are minimal and between-groups differences are maximized. Therefore, LCA can adequately deal with the issue of heterogeneity since the within-group class shares the similar risk preference patterns. Second, followed the methodologies provided by previous papers (see Finkelstein and Poterba [14] Finkelstein and McGrary [13] Chiappori et al. [7]), we use a bivariate probit model that estimate the two probit equations simultaneously to implement the standard positive correlation test for asymmetric information.

Table 3: Bivariate probit regression results of the health insurance demand.

Variables	**Model 1**	**Model 2**	**Model 3**
Latent Class Classification			
Class 2	-0.1039	-0.1076	-0.1014
	[0.1091]	[0.1091]	[0.1092]
Class 3	-0.1801***	-0.1988***	-0.1999***
	[0.0650]	[0.0653]	[0.0654]
Personal Characteristics			
Age between 40 and 45	0.5704***	0.4924***	0.5004***
	[0.0.0737]	[0.0764]	[0.0765]
Age between 45 and 50	0.5237***	0.4558***	0.4646***
	[0.0729]	[0.0750]	[0.0751]
Age between 50 and 55	0.4060***	0.3578***	0.3616***
	[0.0725]	[0.0738]	[0.0738]
Age between 55 and 60	0.2759***	0.2439***	0.2518***
	[0.0793]	[0.0800]	[0.0801]
Gender/Male	-0.2379***	-0.2262***	-0.2290***
	[0.0413]	[0.0800]	[0.0426]
Marital status	-0.0075	0.0014	-0.0018
	[0.0556]	[0.0557]	[0.0558]
Educational Attainment			
Junior high school	0.0880	0.0824	0.0785
	[0.0565]	[0.0568]	[0.0569]
Senior high school	0.3164***	0.3135***	0.3050***
	[0.0544]	[0.0547]	[0.0548]
College or above	0.4815***	0.4688***	0.4571***
	[0.0665]	[0.0665]	[0.0667]
Regional Variables			
Central area	0.0664	0.0638	0.0637
	[0.0488]	[0.0488]	[0.0490]
Southern area	0.0781*	0.0803*	0.0193*
	[0.0476]	[0.0478]	[0.0478]
Eastern area	0.0242	0.0211	0.0785
	[0.0714]	[0.0717]	[0.0718]
Individual Monthly Income Level			
NT$15,000 to 20,000	0.1732***	0.1786***	0.1708***
	[0.0540]	[0.0542]	[0.0543]
NT$20,000 to 40,000	0.3380***	0.3528***	0.3430***
	[0.0572]	[0.0573]	[0.0574]
NT$40,000 to 60,000	0.5164***	0.5244***	0.5150***
	[0.0638]	[0.0639]	[0.0640]
NT$60,000 to 80,000	0.5826***	0.5945***	0.5857***
	[0.0784]	[0.0783]	[0.0786]
NT$80,000 to100,000	0.7022***	0.7244***	0.7130***
	[0.1350]	[0.1350]	[0.1355]
Above NT$ 100,000	0.5067***	0.4990***	0.5039***
	[0.1324]	[0.1326]	[0.1329]
Ethnicity			
Taiwanese	0.0514	0.0484	0.0459
	[0.0685]	[0.0689]	[0.0689]
Hakka	-0.0332	-0.0391	-0.0404
	[0.0823]	[0.0827]	[0.0827]
Mainlander	-0.4360***	-0.4106**	-0.4149***
	[0.1685]	[0.1691]	[0.1692]

Table 3: (continued)

Variables	**Model 1**	**Model 2**	**Model 3**
Health status			
Good	0.1804		0.1894
	[0.1244]		[0.1248]
Fair	0.2209*		0.2324*
	[0.1221]		[0.1226]
Poor	0.1341		0.1553
	[0.1217]		[0.1225]
Very poor	0.0240		0.0494
	[0.1424]		[0.1456]
Diseases			
Hypertension		-0.2035***	-0.1992***
		[0.0593]	[0.0594]
Diabetes		-0.2332***	-0.2192***
		[0.0761]	[0.0767]
Asthma		-0.1682	-0.1690
		[0.1400]	[0.1399]
Stroke		0.0149	0.0403
		[02331]	[0.2336]
Kidney disease		0.0417	0.0505
		[0.0845]	[0.0846]
Heart disease		0.0771	0.0890
		[0.1038]	[0.1041]
Gout		-0.0843	-0.0788
		[0.0803]	[0.0804]
Stomach disease		0.0097	0.0184
		[0.0766]	[0.0769]
Nose disease		0.1251	0.1312
		[0.0893]	[0.0894]
Lung disease		-0.1343	-0.1250
		[0.1212]	[0.1214]
Osteoporosis		0.0095	0.0165
		[0.0636]	[0.0638]
Liver disease		0.1799***	0.1926***
		[0.0730]	[0.0734]
Cancer		-0.0209	-0.0067
		[0.1688]	[0.1698]
Mental Disorder		0.0422	0.0693
		[0.1134]	[0.1142]
Arthritis		0.0220	0.0384
		[0.0823]	[0.0828]
Instrumentals			
Look	-0.1297***	-0.1334***	-0.1321***
	[0.0407]	[0.0410]	[0.0411]
Ask	-0.0245	-0.0186	0.0205
	[0.0444]	[0.0442]	[0.0441]
Sample Size		4921	

Note1: Significant at 10 significant at 5%; significant at 1%.

Note2: Class 1:“high cautiousness”; Class 2: “low cautiousness without wearing a seat belt or helmet” and Class 3:” low cautiousness with smoking behavior”.

Note3: the default variable is class 1; in age category is age between 60 and 65; in educational attainment is elementary school or below; in regional variable is the northern area; in individual monthly income level is below NT$15,000; in ethnicity is other; in health status is very good.

Table 4: Bivariate probit regression results of the inpatient services utilization.

Variables	Model 1	Model 2	Model 3
Latent Class Classification			
Class 2	0.1146	0.1419	0.1364
	[0.1480]	[0.1509]	[0.1505]
Class 3	-0.0996	-0.0650	-0.0681
	[0.0990]	[0.1006]	[0.1005]
Class2*PHI	-0.0717	-0.0459	-0.0571
	[0.2543]	[0.2622]	[0.2594]
Class3*PHI	0.2630*	0.2515*	0.2545*
	[0.1484]	[0.1522]	[0.1505]
Private Health Insurance (PHI)	0.9446**	0.8783**	0.9234***
	[0.3941]	[0.3756]	[0.3474]
Personal Characteristics			
Age between 40 and 45	-0.3600***	-0.2073**	-0.2195**
	[0.0999]	[0.1045]	[0.1025]
Age between 45 and 50	-0.3102***	-0.1767*	-0.1970**
	[0.0952]	[0.0996]	[0.0977]
Age between 50 and 55	-0.2297***	-0.1575*	-0.1699*
	[0.0884]	[0.0912]	[0.0901]
Age between 55 and 60	0.0092	0.0593	0.0479
	[0.0925]	[0.0942]	[0.0935]
Gender/Male	0.2356***	0.2083***	0.2120***
	[0.0556]	[0.0578]	[0.0571]
Marital status	-0.1334**	-0.1407**	-0.1333**
	[0.0668]	[0.0675]	[0.0673]
Educational Attainment			
Junior high school	-0.0839	-0.0933	-0.0942
	[0.0684]	[0.0696]	[0.0695]
Senior high school	-0.2850***	-0.2992***	-0.2929***
	[0.0739]	[0.0750]	[0.0738]
College or above	-0.4518***	-0.4840***	-0.4713***
	[0.0985]	[0.0979]	[0.0958]
Regional Variables			
Central area	0.0469	0.0596	0.0562
	[0.0619]	[0.0635]	[0.0630]
Southern area	0.0219	0.0100	0.0040
	[0.0611]	[0.0621]	[0.0616]
Eastern area	-0.0664	-0.0867	-0.0806
	[0.0931]	[0.0946]	[0.0945]
Individual Monthly Income Level			
NT$15,000 to 20,000	-0.0463	-0.0579	-0.0408
	[0.0671]	[0.0524]	[0.0679]
NT$20,000 to 40,000	-0.2164***	-0.2173***	-0.2002***
	[0.0788]	[0.0801]	[0.0792]
NT$40,000 to 60,000	-0.2732***	-0.2763***	-0.2605***
	[0.1009]	[0.1005]	[0.0980]
NT$60,000 to 80,000	-0.1418	-0.1462	-0.1250
	[0.1307]	[0.1277]	[0.1249]
NT$80,000 to100,000	-0.3420*	-0.3389*	-0.3104
	[0.1977]	[0.1991]	[0.1965]
Above NT$ 100,000	-0.2532	-0.2221	-0.2053
	[0.1872]	[0.1881]	[0.1873]

Table 4: (continued)

Variables	Model 1	Model 2	Model 3
Ethnicity			
Taiwanese	-0.1231	-0.0970	-0.1036
	[0.0878]	[0.0904]	[0.0900]
Hakka	-0.1410	-0.1274	-0.1379
	[0.1064]	[0.1093]	[0.1090]
Mainlander	0.2173	0.1788	0.1688
	[0.1911]	[0.1963]	[0.1964]
Health status			
Good	0.0614		0.0364
	[0.1831]		[0.1849]
Fair	0.1318		0.0877
	[0.1820]		[0.1829]
Poor	0.2184		0.1220
	[0.1798]		[0.1811]
Very poor	0.6546***		0.3996**
	[0.1984]		[0.1980]
Diseases			
Hypertension		0.1920***	0.1817***
		[0.0686]	[0.0684]
Diabetes		0.1908**	0.1661**
		[0.0858]	[0.0856]
Asthma		0.3856***	0.3907***
		[0.1411]	[0.1406]
Stroke		0.1603	0.1048
		[0.2376]	[0.2373]
Kidney disease		0.1012	0.0764
		[0.0983]	[0.0984]
Heart disease		0.2105*	0.1781
		[0.1175]	[0.1170]
Gout		0.2184**	0.2020**
		[0.0911]	[0.0911]
Stomach disease		0.2099**	0.1820**
		[0.0886]	[0.0885]
Nose disease		-0.1884	-0.1917
		[0.1217]	[0.1209]
Lung disease		-0.0880	-0.1077
		[0.1436]	[0.1430]
Osteoporosis		0.2190***	0.1959***
		[0.0765]	[0.0761]
Liver disease		0.2643***	0.2277***
		[0.0935]	[0.0922]
Cancer		0.6795***	0.6176***
		[0.1727]	[0.1716]
Mental Disorder		0.3348***	0.32758**
		[0.1280]	[0.1282]
Arthritis		0.0499	0.0128
		[0.0950]	[0.0955]
Sample Size		4921	

Note1: Significant at 10%; significant at 5%; significant at 1%.

Note2: Class 1:"high cautiousness"; Class 2: "low cautiousness without wearing a seat belt or helmet" and Class 3:" low cautiousness with smoking behavior".

Note3: The default variable is class 1; in age category is age between 60 and 65; in educational attainment is elementary school or below; in regional variable is the northern area; in individual monthly income level is below NT$15,000; in ethnicity is other; in health status is very good.

Our results of the negatively significant correlation coefficients ρ that violates the positive correlation in all models may indicate advantageous selection exist in the supplementary health insurance market. Since relative to the high cautiousness class, individuals in the low cautiousness with smoking behavior class tend to have lower probability to buy PHI and have higher probability to utilization inpatient service in the subsequent 2 years if insured in all models. In other words, individuals in the high cautiousness class may suggest the existence of advantageous selection because they tend to have higher probability to buy PHI and have lower probability to utilization inpatient service in the subsequent 2 years if insured. Our results are similar to previous findings such as Fang et al. [12]. They found that advantageous selection exists in Medigap market.

In conclusion, previous literature had showed the heterogeneity in risk preferences is an important factor in determining individuals insurance purchase decision. In this paper, we take unobserved heterogeneity of individuals risk preferences into account by adopting LCA. We further examine the relationship between unobserved risk preferences and insurance purchase decision and the subsequent inpatient service utilization. As a result, our study is the first research that not only uses LCA to examine the relationship between risk preference heterogeneity and purchase of PHI but also the relationship between risk preference heterogeneity and the subsequent inpatient service utilization.

Our results shed some light on the possible factors of health insurance claim probability. Specifically, it is worth noting that Taiwan has implemented the NHI program since 1995. The NHI is a universal and compulsory social health insurance program (the coverage rate is 99% above). PHI in Taiwan is regarded as a supplementary insurance to the public health system (see Tian et al. [23]). Our empirical findings provide evidences regarding asymmetric information of the supplementary health insurance market. If the insurance company does not consider the risk preference in insurance pricing, they could have mispricing the insurance. Our empirical findings could also provide information about how insurance companies can perform a target market strategy on their potential customers.

Acknowledgements

This paper is supported by the Ministry of Science and Technology of Taiwan under grant MOST 102-2410-H-032-031.

References

[1] Aday, L. A. and Andersen, R. (1974). *A framework for the study of access to medical care*, Health Services Research, Vol.9, 208-220.

[2] Alessie, R., Angelini, V., Mierau, J. O. and Viluma, L. (2020). *Moral hazard and selection for voluntary deductibles*, Health Economics, Vol.29, 1251-1269.

[3] Andersen, R. and Newman, J. F. (2005). *Societal and Individual Determinants of Medical Care Utilization in the United States*, The Milbank Quarterly, Vol.83, 95-124.

[4] Bago d'Uva T. (2005). *Latent class models for use of primary care: evidence from a British panel*, Health Economics, Vol.14, 873-892.

[5] Bago d'Uva T. (2006). *Latent class models for utilisation of health care*, Health Economics, Vol.15, 329-343.

[6] Cawley, J. and Philipson, T. (1999). *An Empirical Examination of Information Barriers to Trade in Insurance*, The American Economic Review, Vol.89, 827-46.

[7] Chiappori, P., Jullien, B., Salanié, B. and Salanié, F. (2006). *Asymmetric Information in Insurance: General Testable Implications*, The RAND Journal of Economics, Vol.37, 783-798.

[8] Cohen, A. and Siegelman, P. (2010). *Testing for Adverse Selection in Insurance Markets*, The Journal of Risk and Insurance, Vol.77, 39-84.

[9] Cutler, D. M., Finkelstein, A. and McGarry, K. (2008). *Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance*, The American Economic Review, Vol.98, 157-162.

[10] Deb, P. and Trivedi, P. K. (2002). *The structure of demand for health care: latent class versus two-part models*, Journal of Health Economics, Vol.21, 601-625.

[11] Einav, L., Finkelstein, A. and Levin, J. (2010). *Beyond Testing: Empirical Models of Insurance Markets*, Annual Review of Economics, Vol.2, 311-336.

[12] Fang, H., Keane, M. and Silverman, D. (2008). *Sources of Advantageous Selection: Evidence from the Medigap Insurance Market*, Journal of Political Economy, Vol.116, 303-350.

[13] Finkelstein, A., and McGarry, K. (2006). *Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market*, The American Economic Review, Vol.96, 938-958.

[14] Finkelstein, A. and Poterba, J. (2004). *Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market*, Journal of Political Economy, Vol.112, 183-208.

[15] Greene W., Harris, M., Hollingworth, B. and Maitra, P. (2008), *A Bivariate Latent Class Correlated Generalized Ordered Probit Model with an Application to Modeling Observed Obesity Levels*, Working Papers No. 2451/26027, New York University, Leonard N. Stern School of Business, Department of Economics.

[16] Lafortune, L., Béland, F., Bergman, H. and Ankri, J. (2009). *Health state profiles and service utilization in community-living elderly*, Medical Care, Vol.47, 286-294.

[17] Laska, M. N., Pasch, K. E., Lust, K., Story, M. and Ehlinger, E. (2009). *Latent class analysis of lifestyle characteristics and health risk behaviors among college youth*, Prevention Science, Vol.10, 376-386.

[18] Li Donni, P. (2010). *Risk preference heterogeneity and multiple demand for insurance, Health*, Econometrics and Data Group (HEDG) Working Papers 10/17, HEDG, c/o Department of Economics, University of York.

[19] Puelz, R., and Snow, A. (1994). *Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market*, Journal of Political Economy, Vol.102, 236-257.

[20] Rothschild, M., and Stiglitz, J. (1976). *Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*, The Quarterly Journal of Economics, Vol.90, 629-649.

[21] Scharoun-Lee, M., Kaufman, J. S., Popkin, B. M. and Gordon-Larsen, P. (2009). *Obesity, race/ethnicity and life course socioeconomic status across the transition from adolescence to adulthood*, Journal of Epidemiology and Community Health, Vol.63, 133-139.

[22] Sloan, F. A., Robinson, P. A., and Eldred, L. M. (2018). *Advantageous Selection, Moral Hazard, and Insurer Sorting on Risk in the U.S. Automobile Insurance Market*, The Journal of Risk and Insurance, Vol.85, 545-575.

[23] Tian, W. H., Tien, J.T., Chen, C.S. and Liu, T.C. (2012). *Private Health Insurance and Inpatient Service Utilisation among Adults and Elderly People under Taiwan's National Insurance Programme*, Geneva Risk and Insurance Issue and Practice, Vol 37, 655-677.

[24] Wu, S. (2003). *Sickness and preventive medical behavior*, Journal of Health Economics, Vol.22, 675-689.

Department of Risk Management and Insurance, Tamkang University, Taiwan, ROC.

E-mail: jilltien@mail.tku.edu.tw

Major area (s): Risk evaluation and management, health insurance, ageing society issue.

Department of Economics, National Cheng Kung University, Taiwan, ROC.

E-mail: whtian@mail.ncku.edu.tw

Major area (s): Health policy, health economics, labor economics.

(Received March 2021; accepted July 2021)