

EVIDENCE OF ASYMMETRIC INFORMATION IN THE AUTOMOBILE INSURANCE MARKET: DICHOTOMOUS VERSUS MULTINOMIAL MEASUREMENT OF INSURANCE COVERAGE

Hyojoung Kim
Doyoung Kim
Subin Im
James W. Hardin

ABSTRACT

In the empirical analysis of information asymmetry in automobile insurance markets, prior research used a dichotomous measurement approach that induces excessive bundling in coverage measurements and sample selection biases. To improve on the conditional correlation method for testing information asymmetry, we propose a multinomial measurement approach that constructs coverage categories at ordered multinomial levels. With this approach, we find robust evidence of information asymmetry in both coverage area and coverage amount choices, which we could not find with the dichotomous measurement approach. It thus demonstrates the sensitivity of the empirical findings to the method used to measure insurance coverage.

INTRODUCTION

Since Akerlof (1970), “information asymmetry” has become one of the most influential theoretical concepts in explaining market dynamics. In a nutshell, asymmetric information is problematic because it creates opportunities for lying and cheating by agents, which can induce market inefficiency or even failure. In insurance markets, information asymmetry arises when an informational gap exists between an insured

Hyojoung Kim is at the Department of Sociology, California State University, Los Angeles. Doyoung Kim is at the Department of Economics, Sogang University, Korea. Subin Im is at the Department of Marketing, San Francisco State University. James W. Hardin is at the Department of Epidemiology and Biostatistics, University of South Carolina. Hyojoung Kim can be contacted via e-mail: hkim@calstatela.edu or Doyoung Kim can be contacted via e-mail: dkim@sogang.ac.kr. We thank the insurance company in Korea that graciously allowed us to use its insurance contract data. Regretfully, we cannot reveal its name to preserve confidentiality. We also thank Georges Dionne and two anonymous referees of JRI for their helpful comments. Doyoung Kim acknowledges financial support from the Sogang University Foundation Research Grants in 2008.

and an insurer, in that the latter has less information about the former's risk factors. The resultant asymmetry in risk assessments creates opportunities for strategic behaviors such as adverse selection and moral hazard (Rothschild and Stiglitz, 1976; Arnott and Stiglitz, 1988).

Much effort has been expended on empirically documenting the presence of information asymmetry in various insurance markets. Although studies show that information asymmetry exists in many insurance markets, such as health and life insurance,¹ "the evidence is non-conclusive on the existence of residual asymmetric information in automobile insurance markets" (Dionne, Michaud, and Dahchour, 2006, p. 2). In the past, the vast majority of prior research found no evidence of it in automobile insurance markets, such as those in France, Canada, and Japan (Chiappori and Salanié, 1997, 2000; Richaudeau, 1999; Dionne, Gouriéroux, and Vanasse, 2001; Saito, 2006).² These studies have led to a widespread perception that there is no information asymmetry in the automobile insurance market.

Recently, however, this perception is being challenged. For example, Cohen (2005) finds evidence for information asymmetry in the Israeli automobile insurance market. Whereas this finding is limited in its applicability because the Israeli market is unique in that it lacks an information sharing system in which insurance companies share information about policyholders' risk factors, other studies proffer evidence for moral hazard with broader applicability. Dionne, Michaud, and Dahchour (2006) analyze 3-year panel data in France and find robust evidence for moral hazard especially among policyholders of 5–15 years of driving history. In a similar vein, Li, Liu, and Yeh (2007) and Wang, Chung, and Tzeng (2008) also report evidence for moral hazard while utilizing the unique nature of the increasing deductibles scheme in Taiwan.³

A careful examination suggests that the divide in the literature largely coincides with the nature of the data utilized. Setting aside the Israel case with no information sharing system among insurers, the studies that find evidence for moral hazard utilize either a "natural experiment" attribute of the Taiwanese increasing deductibles scheme or multiyear dynamic panel data. In contrast, almost all the studies that report no evidence of information asymmetry use standard static cross-sectional data.

Against this background, we notice a potential problem in the conventional measurement approach applied to test information asymmetry using cross-sectional data. That is, prior research reduces the multinomial choice categories to more limited, dichotomous choice categories when measuring insurance coverage. To take the French automobile insurance market as an example, French drivers typically receive, in addition to compulsory liability insurance (RC contract), the option to choose collision

¹ See Chiappori, Durand, and Geoffard (1998), Holly, Gardiol, Domenighetti, and Bisig (1998), Cardon and Hendel (2001), and Vera-Hernandez (2003) for evidence of information asymmetry in health insurance; Hendel and Lizzeri (2003) for life insurance; and Finkelstein and Poterba (2004) for annuity contracts.

² Although a pioneering study by Puelz and Snow (1994) reports evidence of information asymmetry in the American market, their finding has been largely discredited due to suspected model misspecifications (Dionne, Gouriéroux, and Vanasse, 2001).

³ For evidence of *ex post* moral hazard (such as fraudulent claims making) in automobile insurance markets, see Cummins and Tennyson (1996) and Dionne and Gagné (2002).

and personal injury protection. Chiappori and Salanié (2000) apparently collapse these two optional coverage areas into a TR contract and measure insurance coverage by pitting the RC contract against the TR contract. In a slightly different manner, Richaudeau (1999) purposely ignores the personal injury protection option and focuses on the choice between liability and collision, unwittingly bundling those who purchase personal injury protection into the group of collision or the group of liability only (depending on whether they purchase collision or not). As Chiappori and Salanié recognize, this dichotomous measurement approach is likely to induce a problem of excessive bundling in measuring insurance coverage.

In addition, this approach can introduce a sample selection bias. In analyzing information asymmetry in the choice of coverage amounts, such as collision deductibles, researchers often combine coverage amounts into dichotomous categories of “small” and “large” (e.g., Puelz and Snow, 1994; Cohen, 2005). In the process, they exclude those who opt not to purchase any collision coverage.⁴ As long as policyholders have the option to not buy collision insurance, the choice of *no* collision insurance must be treated as the choice of an *infinite* collision deductible.

To reduce the extent of bundling in measuring coverage choices and remove the accompanying selection bias, we propose an alternative measurement approach. We view insurance coverage as consisting of two dimensions, that is, coverage area and coverage amount within each area. Accordingly, we construct an ordered multinomial scale of coverage while focusing on each dimension separately. For example, we assess a policyholder’s choice of coverage areas along an ordered multinomial scale of “none,” “some,” and “all” optional coverage areas. For the coverage amount dimension, we treat the choice of no coverage as either an “infinite” deductible (in the case of collision) or “zero” coverage (for the other coverage areas). Accordingly, we measure a policyholder’s choice of collision deductibles along another ordered multinomial scale of “infinite,” “large,” and “small” deductibles. The multinomial measurement approach that we propose can replicate the choice situations that a policyholder faces in practice without excluding anyone from the analysis. In addition, it can pinpoint which choice of insurance coverage (e.g., the choice between none and some, the choice between some and all optional coverage areas, or both) becomes particularly subject to policyholders’ opportunism under asymmetric information. Therefore, our proposed approach helps overcome the limitations of dichotomous measures and provides a more accurate and comprehensive test of the theoretical predictions.

Applying this multinomial measurement approach, we examine the automobile insurance market in South Korea (hereafter, “Korea”). The Korean automobile insurance market is comparable to other advanced markets; as in the United States, Canada, and France, the Korean market has an information sharing system that offers insurance companies free and full access to a national database of the complete records of policyholders on claims for accidents in which they have been involved. We find that the conventional dichotomous measurement approach fails to produce evidence of information asymmetry, either in coverage area or coverage amount choices. Yet with the multinomial measurement approach, we uncover robust evidence for

⁴ In our sample, 49 percent of policyholders opt not to purchase collision coverage.

information asymmetry in coverage area choices as well as coverage amount choices in the areas of collision and personal injury protection. These findings also demonstrate the limitations of the conventional dichotomous measures.

The rest of the article proceeds as follows: In the next section, we present our proposed multinomial measurement method and specify the econometric models to test for information asymmetry. The data section discusses the automobile insurance market in Korea and the data we use for our analysis. After reporting the findings in results section, we conclude the paper by discussing the implications of our findings.

COVERAGE MEASUREMENT AND CONDITIONAL CORRELATION APPROACH

Multinomial Measurement of Insurance Coverage

In most automobile insurance markets, insurance companies offer, in addition to legally mandated liability coverage, a menu of optional coverage areas. For each coverage area, a policyholder also chooses among different amounts of coverage. Note that coverage areas differ from each other in kind, making it difficult to summarize the overall coverage level in a single measure. For example, we cannot directly compare collision coverage with a deductible of \$1,000 with personal injury protection coverage of \$1,000. The challenge therefore is how to categorize coverage levels meaningfully across heterogeneous coverage areas without committing excessive bundling.

Accordingly, we measure insurance coverage as follows. We first assume that insurance coverage consists of two dimensions: the areas covered and coverage amount for each area. We measure each coverage dimension—that is, *coverage area* and *coverage amount* within each coverage area—by constructing sets of three rank-ordered categories. For coverage area, these ordered categories are “no” optional coverage, “some” optional coverage, or “all” optional coverage.⁵ In this scheme, the legally mandated minimum insurance coverage (liability) constitutes the lowest limit (no optional coverage) of insurance coverage. Within each coverage area, we construct the following set of three coverage amount categories in rank order: “zero,” “small,” and “large.” The lower limit of a zero coverage amount refers to those who do not purchase the corresponding coverage area. For others, the coverage amount may be either small or large, depending on the coverage amount they buy. Thus, our measurement scheme leaves out no policyholders in the sample, including those who opt not to purchase a coverage area.

This multinomial measurement approach provides several advantages over the methods used by previous studies. First, by including all policyholders in the sample, it eliminates the possibility of a selection bias. Second, it helps close the gap between real versus measured coverage choices and thus reduces the risk of excessive bundling. Third, by examining the entire insurance package, our measurement approach can not only indicate whether information asymmetry exists, but also pinpoint which aspects of insurance coverage are subject to the predicted behavioral consequences of information asymmetry.

⁵ A simple count of coverage areas would not help because each coverage area differs from the others in kind.

Conditional Correlation Between Coverage and Accidents

With information asymmetry, an agent of a higher risk type chooses a contract with more comprehensive coverage than does a lower-risk agent, a behavioral consequence known as adverse selection (Rothschild and Stiglitz, 1976). This prediction implies a positive correlation between coverage and accidents for observationally equivalent policyholders. Put differently, there is a positive correlation between coverage and accidents, conditional on observable symmetric information in insurance contracts.

Chiappori et al. (2006) demonstrate that the positive correlation property is remarkably robust in diverse situations and therefore can be adopted as the standard procedure for testing information asymmetry and consequent adverse selection. In particular, they show that in the case of competitive insurance markets the positive correlation property is preserved even with heterogeneous preferences, multiple loss levels, and multidimensional adverse selection with possible moral hazard. In the case of imperfectly competitive markets, the positive correlation property still holds unless an agent has private information about his or her risk aversion.⁶

The conditional correlation approach, however, has a limit. As Arnott and Stiglitz (1988) and others point out, information asymmetry gives rise to not only adverse selection but also moral hazard. For example, a better insured agent, given this extensive protection, has less incentive to take preventive measures and may tend to cause more accidents than he or she would with less coverage. Adverse selection thus gets compounded by moral hazard, which also suggests a positive correlation between insurance coverage and accidents. It follows that in a static cross-sectional context, the conditional correlation approach does not help identify the precise type of information asymmetry that underlies the observed correlation.⁷ For this reason, Chiappori and Salanié (2000) call for caution in interpreting test outcomes: although a lack of a conditional correlation provides conclusive evidence of the absence of adverse selection, its presence does not necessarily imply adverse selection.

Accordingly, a positive correlation between insurance coverage and accidents, as we document herein, can indicate the presence of information asymmetry, yet we are still left uncertain about which behavioral consequence it has prompted, namely, adverse selection, moral hazard, or both.

⁶ Chiappori et al. (2006) also show special cases in which information asymmetry may imply a negative correlation (see also de Meza and Webb 2001; Jullien, Salanié, and Salanié 2007). For instance, with the introduction of an administrative cost, if an agent's risk is perfectly correlated with his or her risk aversion in that more risk-averse agents are less risky, asymmetric information about risk and risk aversion may imply a negative correlation between coverage and accidents. Similarly, if asymmetric information about risk aversion is combined with endogenous moral hazard in a way that more risk-averse agents choose more preventive effort, a negative correlation can result. But these are somewhat extreme cases as they require imperfect competition, specific relation between risk aversion and risk, and others.

⁷ According to Chiappori (2000), researchers can test empirically for moral hazard in a "natural experiment" context in which a change occurs exogenously in agents' incentive structure (as in Chiappori, Durand, and Geoffard, 1998), or if observationally equivalent agents are offered different incentive schemes (as in Cardon and Hendel, 2001), or by investigating the dynamics of the conditional correlations (as in Abbring, Chiappori, and Pinquet, 2003).

Econometric Models

Briefly, the conditional correlation approach builds on the following two-stage regression analysis (Richaudeau, 1999). In the first stage, we regress a policyholder's coverage choice on a set of independent variables observable to the insurance company. From this regression, we can estimate a measure (generalized residuals) of the unobservable hidden information possessed by the policyholder when making a coverage choice. In the second stage, we regress actual accident occurrences on the same set of observables and on the estimated coverage choice residuals obtained from the first-stage regression. The coefficients of the coverage choice residuals in the latter regression reveal the conditional correlation between coverage and accidents and thus provide test results for asymmetric information.⁸ The first-stage coverage choice equation gives the statistical inference of the theoretical construct of hidden information, and therefore its model specification is subject to modifications according to the different measurement approaches discussed previously.

Coverage Choice Equation. To specify the first-stage coverage choice equation, let C_i refer to a policyholder i 's coverage choice. His or her choice occurs at the beginning of a contract year. The generic form of a contract choice equation is

$$C_i = f(X_i) + \varepsilon_i, \quad (1)$$

where ε_i is a usual econometric error term and X_i refers to, ideally, all observables of an insurance company (public information).

The nature of C_i dictates which estimation method to use. We first explain the econometric model, which measures C_i dichotomously. For notational convenience, let C_i^b refer to this binary choice dependent variable. For example, C_i^b may take the form of a choice between liability only ($C_i^b = 0$) versus additional collision ($C_i^b = 1$), as in Chiappori and Salanié (1997, 2000) and Richaudeau (1999). Replacing C_i in Equation (1) with C_i^b , we estimate the resultant equation by a maximum-likelihood probit regression and obtain the following generalized residual:

$$\hat{\varepsilon}_i^b = \frac{\phi(X_i \hat{\alpha}_b)}{\Phi(X_i \hat{\alpha}_b)(1 - \Phi(X_i \hat{\alpha}_b))} [C_i^b - \Phi(X_i \hat{\alpha}_b)], \quad (2)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of the standard normal distribution, respectively, and $\hat{\alpha}_b$ is an estimated coefficient vector. As an unexplained probability of making a corresponding coverage choice, $\hat{\varepsilon}_i^b$ effectively captures the extent of private information in the binary choice of C_i , conditional on the observables (Richaudeau, 1999).

When C_i in Equation (1) represents an ordered multinomial contract choice, we let C_i^m refer to the ordered multinomial choice dependent variable and code C_i^m as 0 if a

⁸ Similarly, Chiappori and Salanié (1997, 2000) regress the coverage choice and accident occurrence equations on a common set of observable variables, then compute the conditional correlation using the residuals of each regression. However, this approach is somewhat restrictive because it requires the same distribution for both equations.

policyholder i chooses no optional coverage (i.e., liability only), 1 for some optional coverage, and 2 for all optional coverage. We replace C_i in Equation (1) with C_i^m to estimate the multinomial choice model:

$$C_i^m = \begin{cases} 0 & \text{if } C_i^{m*} \leq \mu_1 \\ 1 & \text{if } \mu_1 < C_i^{m*} \leq \mu_2 \\ 2 & \text{if } \mu_2 < C_i^{m*} \end{cases} \quad (3)$$

where C_i^{m*} is a latent variable representing the policyholder's utility associated with insurance coverage, and μ_1 and μ_2 are unknown thresholds for observed categories. We estimate this ordered multinomial choice model using an ordered probit regression.

If the coverage choice dependent variable is binary, the unexplained probabilities, which represent an estimate of asymmetric information in the coverage choice, can be obtained easily by subtracting the predicted probabilities from the observed binary choices of $C_i^b \in \{0, 1\}$. However, with an ordered multinomial dependent variable of $C_i^m \in \{0, 1, 2\}$, this procedure is not applicable. To obtain the unexplained probabilities equivalent to the one in the binary choice model, we group three choices of $C_i^m \in \{0, 1, 2\}$ into two ordered choice sets. In particular, let $C_i^1 = 0$ if $C_i^m = 0$ and $C_i^1 = 1$ if $C_i^m \in \{1, 2\}$. Also let $C_i^2 = 0$ if $C_i^m \in \{0, 1\}$ and $C_i^2 = 1$ if $C_i^m = 2$. We then calculate the unexplained probabilities associated with C_i^1 and C_i^2 . That is, the generalized residuals, equivalent to that in Equation (2), are calculated by

$$\hat{\varepsilon}_i^{m1} = \frac{\phi(X_i \hat{\alpha}_m - \hat{\mu}_1)}{\Phi(X_i \hat{\alpha}_m - \hat{\mu}_1)(1 - \Phi(X_i \hat{\alpha}_m - \hat{\mu}_1))} [C_i^1 - \Phi(X_i \hat{\alpha}_m - \hat{\mu}_1)] \quad (4)$$

$$\hat{\varepsilon}_i^{m2} = \frac{\phi(X_i \hat{\alpha}_m - \hat{\mu}_2)}{\Phi(X_i \hat{\alpha}_m - \hat{\mu}_2)(1 - \Phi(X_i \hat{\alpha}_m - \hat{\mu}_2))} [C_i^2 - \Phi(X_i \hat{\alpha}_m - \hat{\mu}_2)], \quad (5)$$

where $\hat{\alpha}_m$ is an estimated coefficient vector, and $\hat{\mu}_1$ and $\hat{\mu}_2$ are estimated thresholds for observed categories. Consequently, $\hat{\varepsilon}_i^{m1}$ estimates the unexplained probability of choosing *any* optional coverage ($C_i^m = \{1, 2\}$) over *no* optional coverage ($C_i^m = 0$), whereas $\hat{\varepsilon}_i^{m2}$ estimates the unexplained probability of choosing *all* optional coverage areas ($C_i^m = 2$) over *fewer than all* optional coverage areas ($C_i^m = \{0, 1\}$). When we include these two residuals in an accident occurrence equation (see Equation (6)) as regressors, the estimated regression coefficient of $\hat{\varepsilon}_i^{m1}$ can capture the effect of information asymmetry on the choice between no optional coverage ($C_i^m = 0$) and some optional coverage ($C_i^m = 1$), and that of $\hat{\varepsilon}_i^{m2}$ can capture the effect of information asymmetry on the choice between some optional coverage ($C_i^m = 1$) and all optional coverage ($C_i^m = 2$). We apply the same analytic strategy to the choice of coverage amounts within a coverage area, another dimension of insurance coverage.

Accident Occurrence Equation. Using the residual obtained from the binary choice model in Equation (2) and the residuals from the multinomial choice model in

Equations (4) and (5), it is straightforward to test for information asymmetry. Let A_i refer to a variable that measures the accidents a policyholder i has had during the contract year. The generic form of accident equations can be summarized as follows:

$$A_i = g(X_i, \hat{\varepsilon}_i) + v_i, \quad (6)$$

where v_i is an error term, $\hat{\varepsilon}_i$ refers to the extent of asymmetric information as estimated in Equation (2) for a binary choice model ($\hat{\varepsilon}_i^b$) or in Equations (4) and (5) for a multinomial choice model ($\hat{\varepsilon}_i^{m1}$ and $\hat{\varepsilon}_i^{m2}$), and X_i is a vector of the same observables (public information) used in the coverage choice equation (Equation (1)).

The exact nature of the g function depends on the measurement of A_i . Because we measure A_i as the total number of accidents that a policyholder i has caused during the contract year, we estimate Equation (6) using a negative binomial regression,⁹ as follows:

$$P(A_i) = \frac{\Gamma\left(A_i + \frac{1}{\sigma^2}\right) [\sigma^2 \exp(X_i \hat{\beta}_0 + \hat{\varepsilon}_i \hat{\beta}_\varepsilon)]^{A_i}}{\Gamma\left(\frac{1}{\sigma^2}\right) \Gamma(A_i + 1) [1 + \sigma^2 \exp(X_i \hat{\beta}_0 + \hat{\varepsilon}_i \hat{\beta}_\varepsilon)]^{A_i + \frac{1}{\sigma^2}}}, \quad (7)$$

where $\Gamma(\cdot)$ is the Gamma function, and $\hat{\beta}_0$ and $\hat{\beta}_\varepsilon$ are estimated coefficient vectors. The regression coefficient $\hat{\beta}_\varepsilon$ captures the correlation between insurance coverage and accidents, conditional on the effects of all observables on both the coverage choice and subsequent accidents. Thus, it provides the focal test statistics for information asymmetry. In other words, statistically significant and positive $\hat{\beta}_\varepsilon$ indicates the existence of information asymmetry.

Finally, the test procedure involves a two-stage, nonlinear estimation that likely induces biases in the estimated standard errors in the second stage (Murphy and Topel, 1985). We correct for this bias by testing the statistical significance of $\hat{\beta}_\varepsilon$ using the Murphy–Topel standard error estimates (Murphy and Topel, 1985; Hardin, 2002). The Murphy–Topel estimates are applicable when the first-stage coverage choice equation involves a binary dependent variable and a single predicted value appears in the second stage (Richaudeau, 1999; Hardin, 2002). However, when the first-stage coverage choice equation involves a multinomial dependent variable and multiple residual variables are used in the second stage, as in our analysis, the Murphy–Topel estimates must be modified accordingly. In the Appendix, we provide a detailed description of our modification.

DATA

Automobile Insurance Market in Korea

Korean law requires a minimum amount of liability insurance for bodily injury damage to third parties. In addition, insurance companies are legally bound to offer the following optional coverage areas: additional bodily injury coverage, property damage

⁹ The negative binomial model fits our data better than the Poisson model (see statistics for dispersion parameters in Tables 3 and 4). See also Dionne and Vanasse (1992) who show a similar result for Canadian data.

coverage, collision, personal injury protection, and uninsured/underinsured motorist protection coverage. The first coverage option is an extension of the area covered by the mandated minimum liability insurance, the second coverage option covers property damage to third parties, the others cover personal injury and damage to the policyholder's car. A policyholder may choose any combination of these coverage options.

For each of the coverage options, insurance companies independently determine the array of coverage amounts offered. For example, the insurance company in our study offers six different deductible levels for collision coverage. For each contract, it determines the premium by, among other things, the base premium and the bonus-malus rate.¹⁰ Insurance companies determine both the base premium and the bonus-malus rate according to their own marketing strategies. The insurance company we study determines the base premium by risk classification on the basis of the characteristics of the policyholder (e.g., age, gender) and the automobile to be insured (e.g., car size, age). Furthermore, the government-funded Korea Insurance Development Institute centrally compiles policyholders' records of accidents and bonus-malus rates and makes them readily accessible to insurance companies.

Data and Variables

The data we use in our analysis include a random sample of all the contracts of a major insurance company in Korea.¹¹ Using two calendar years of data, from 2000 to 2001, we construct a cross-sectional data set that contains information about each contract in the sample for a full contract year, for a total of 28,689 annual contracts.

At the time of data collection, the studied company offered a single fixed coverage amount for the coverage area of uninsured/underinsured motorist protection; but, for all other optional areas, it provided variable coverage amounts. Specifically, six different deductible levels were on the offer for collision coverage: 50, 100, 200, 300, 500, and 600 thousand won—in U.S. dollars, approximately \$40 (0.4 percent of per capita income) to \$476 (5 percent of per capita income).¹² For personal injury protection, eight different levels of coverage were available: 15, 30, 50, 60, 100, 200, and 400 million won (approximately \$11,905 to \$317,460), as well as infinite coverage. Available for property damage were 20, 30, 40, 50, 60, and 100 million won (approximately \$15,873 to \$79,365), as well as infinite coverage. The company also offered four different levels of additional bodily injury coverage: 100, 200, and 300 million won (approximately \$79,365 to \$238,095), plus infinite coverage. Policyholders may choose any combination of optional coverage areas. We summarize the number of policyholders who chose each coverage area and, within each area, the number who chose each amount in Table 1.

¹⁰ Other adjustment factors include special discount rates for government-owned automobiles, years of driving as determined by the total number of years insured (penalties accruing for less than 3 years), traffic violation records, and other special contract items such as limitations on eligible drivers for the automobile in question.

¹¹ As we show in Table 2, policyholders of this company represent all residential areas in Korea and are diverse in age, car size, car make, and so on.

¹² The average exchange rate was 1,260 won per \$1, and per capita income was \$9,770 in 2000 (Bank of Korea Economic Statistics System; <http://ecos.bok.or.kr>).

TABLE 1
The Number of Policyholders and the Number of Policyholders With Accidents for Each Coverage Area and Corresponding Coverage Amount

Coverage Area	Number of Policyholders for Each Coverage Area	Number of Policyholders for Each Coverage Amount					
Additional bodily injury	Coverage amount	100	200	300	Infinite		
	Policyholders	24,901	4	1	24,662		
Property damage	Policyholders with accidents	2,199 (8.8%)	0 (0%)	0 (0%)	2,173 (8.8%)		
	Coverage amount	20	30	40	50	60	Infinite
	Policyholders	24,881	20,923	3,200	205	92	22
	Policyholders with accidents	2,193 (8.8%)	1,854 (8.9%)	277 (8.7%)	19 (9.3%)	11 (12.0%)	5 (22.7%)
Collision deductible	Coverage amount	0.05	0.1	0.2	0.3	0.5	0.6
	Policyholders	14,734	10,724	426	44	233	30
Personal injury protection	Policyholders with accidents	1,504 (10.2%)	1,078 (10.1%)	47 (11.0%)	8 (18.2%)	29 (12.5%)	1 (33.3%)
	Coverage amount	15	30	50	60	100	400
	Policyholders	23,944	14,295	4,769	72	51	373
	Policyholders with accidents	2,138 (8.9%)	1,257 (8.8%)	429 (9.0%)	6 (8.3%)	308 (9.3%)	103 (21.4%)
Uninsured motorist protection	Policyholders	22,019					
	Policyholders with accidents	2,003 (9.1%)					
Liability only	Policyholders	3,784					
	Policyholders with accidents	55 (1.45%)					

Notes: The total number of policyholders in the sample is 28,689. The number of policyholders for each coverage area includes observations of missing values in coverage amount for the area. There are 230 missing observations of coverage amount for additional bodily injury coverage, 2 for property damage coverage, 3,274 for collision coverage, and 2 for personal injury protection coverage. Coverage amount is stated in million *won* (the average exchange rate was 1,260 won per \$1 in 2000). The percentage of policyholders with bilateral accidents for each coverage area or for each coverage amount is in parentheses.

In our sample of 28,689 contracts, 13.2 percent (3,784 contracts) had liability only (i.e., no optional coverage), 35.7 percent (10,234 contracts) purchased some optional coverage areas, and the remaining 51.1 percent (14,671 contracts) purchased all optional coverage areas. Of the entire sample, 86.8 percent had additional bodily injury coverage, 86.7 percent had property damage coverage, 51.4 percent had collision, 83.5 percent had personal injury protection, and 76.8 percent had uninsured/underinsured motorist protection coverage. With regard to coverage amount, virtually all (99.9 percent) policyholders with additional bodily injury coverage chose infinite coverage. In contrast, 84.1 percent of those with property damage coverage selected the minimum coverage amount. For collision coverage, 93.6 percent chose the minimum deductible (i.e., maximum coverage amount). For personal injury protection, coverage varies, though a majority (59.7 percent) chose the minimum amount.

We use the number of accidents occurred during the contract year as a measure of *ex post* risk.¹³ Because we obtained our data from an insurance company that does not have records of "actual" accidents, our measure of accident occurrences is based on the number of "claimed" accidents only.¹⁴ There is a problem when using claims instead of actual accidents. A policyholder may or may not report an accident depending on the current insurance contract. For instance, a policyholder with liability insurance is not required to report unilateral accidents that inflict damage on his/her car only whereas a policyholder with collision can report such accidents. Furthermore, even if he/she can report, a policyholder may find it in his or her best interest not to report accidents. As in Chiappori and Salanié (2000), we control these problems by restricting our analysis to "bilateral" accidents that involve a third party. Given that only bilateral accidents are considered, the underreporting of accidents is mostly negligible.¹⁵ In our sample, 7.86 percent (2,255 contracts) have had at least one such bilateral accident during the contract year, and of these, 142 contracts claimed two accidents, and 16 contracts made claims on three accidents. Table 1 also reports the number of policyholders with bilateral accidents for each coverage area and for each coverage amount.

Our data include a complete set of information that the insurance company collected about each contract; thus, these data include details about not only the policyholder's coverage and accident claims but also the policyholder's personal characteristics and information about the covered automobile. The data also include the premium and bonus-malus rate. Therefore, our data comprise all the observable variables that the insurance company acquires at the time of contract, which enables us to

¹³ In addition to the number of accidents, it would be interesting to investigate the evidence of information asymmetry using the size or amount of the loss associated with accidents. We leave such an investigation for another article as it is beyond the scope of this article.

¹⁴ Richaudeau (1999) reports no statistically significant difference in the conditional correlations when accident occurrences are measured on the basis of reported accidents or all accidents, including unreported ones.

¹⁵ The decision to report accidents is known to be contaminated by *ex post* moral hazard (Chiappori and Salanié, 2000). We replicated the analysis while using accidents including unilateral ones. The outcomes were qualitatively identical with those reported in the article. It indicates that *ex post* moral hazard in a reporting decision does not affect the results presented here.

control adequately for potential spuriousness in the conditional correlation between insurance coverage and accidents. For the common control variables used in the first-stage coverage choice and the second-stage accident occurrence regressions, we employ 23 variables, consisting of the policyholder's age, sex, and residence; detailed residential areas (nine residence areas); a car usage indicator; car age, size (three sizes), and make (three makes); the remaining value of the car; the premium; and the bonus-malus rate. Table 2 lists the definitions of these variables, along with basic descriptive statistics.

A special note is needed with regard to insured's past driving history. People may vary in their propensity for accidents, and failing to control for such heterogeneity causes problems in testing asymmetric information (Chiappori and Salanié, 2000). Using data from a random sample of car drivers in Canada, Boyer and Dionne (1989) show in their study of the determination of insurance premiums that past driving history (past accidents, traffic violations, and license suspension) positively affects current propensity for accidents. In most countries, including Korea and the United States, such records are public information known to insurance companies and clearly indicate a person's risk type. Thus, omitting the insured's past driving history implies that corresponding public information about risk is treated as private asymmetric information and therefore could overestimate the level of information asymmetry.

The data we obtained have only the accident records of the current contract year. But it has bonus-malus rates that reflect past accidents. A bonus-malus rate ranges from 0.4 to 2.0, with a starting rate at 1.0. The current bonus-malus rate depends on the individual's prior rate. In the case of bonus rate, for example, having no accident during the last 3 years makes the rate 1.0 if the prior rate is higher than 1.1, and decreases the rate by 0.1 if it is 1.1 or below. The malus rate is determined by the number of accidents and the severity of each accident. Each accident increases the rate at least by 0.05. This way the bonus-malus rate reflects the insured's accident records at least for the past 3 years.

We use this bonus-malus rate as a surrogate for past driving history as recent studies do (Saito, 2006; Li, Liu, and Yeh, 2007). Still, a caution is needed. Because a bonus-malus rate in our study is determined on the basis of reported accidents alone, it may underestimate the degree of risk to the extent of unreported accidents. Furthermore, as shown in Boyer and Dionne (1989), the propensity for accidents is affected not only by the past accident history, but by other indicators of past driving history, such as the number of traffic violations and license suspension. Accordingly, our use of bonus-malus rates in lieu of past accidents and other traffic violations may result in errors in measuring risk classification and thereby overestimate evidence for information asymmetry.

This notwithstanding, it should be noted that this potential measurement problem associated with risk classification does not undermine the main point of our analysis that the multinomial measurement approach of insurance coverage provides an improved statistical tool for testing information asymmetry in comparison to the conventional dichotomous measurement approach. For any measurement errors that

TABLE 2

List of Control Variables Used and Their Descriptive Statistics

-
- Age = the policyholder's age (mean = 41.02, std. dev. = 10.42, min = 19, max = 82)
 - Male = 1 if the policyholder is male (82.35%); 0 if female (17.65%)
 - Urban = 1 if the policyholder has residence in a city (49.14%); 0 otherwise (50.86%)
 - Car usage = 1 if the policyholder specifies the primary use of the insured automobile as for commuting and family use (95.66%); 0 otherwise (i.e., commercial use) (4.34%)
 - Car value = the insurance company's estimate of the remaining value of the insured automobile (in the million won units) (mean = 3.030, std. dev. = 3.700, min = 0, max = 174.9)
 - Car age = the age of the insured automobile (mean = 5.018, std. dev. = 2.473, min = 0, max = 13)
 - Premium = the total premium the policyholder paid for the contract in 10,000 won units (mean = 39.166, std. dev. = 26.652, min = 0.1, max = 1,007.3)
 - Bonus-malus rate = the cumulative bonus-malus coefficient applied to the contract in percent points (mean = 69.882, std. dev. = 24.175, min = 40, max = 200)

Car size variables (reference group = petty cars of engine capacity 1,000 cc or less, 9.09%)

- Small cars = 1 if the insured automobile is of engine capacity from 1,001 cc to 1,500 cc (52.65%); 0 otherwise
- Medium cars = 1 if the insured automobile is of engine capacity from 1,501 cc to 2,000 cc (30.49%); 0 otherwise
- Large cars = 1 if the insured automobile is of engine capacity larger than 2,000 cc (7.77%); 0 otherwise

Car-make variables (reference group = other than three major car makers in Korea, 3.56%)

- Kia = 1 if the insured automobile is made by Kia (24.21%); 0 otherwise
- Daewoo = 1 if the insured automobile is made by Daewoo (25.90%); 0 otherwise
- Hyundai = 1 if the insured automobile is made by Hyundai (46.33%); 0 otherwise

Residence area variables (reference group = Kangwon and Jeju province, 2.98%)

- Seoul = 1 if the residence area of the policyholder is Seoul city (17.24%); 0 otherwise
 - Busan = 1 if the residence area of the policyholder is Busan city (9.65%); 0 otherwise
 - Kyungki = 1 if the residence area of the policyholder is Kyungki province or Incheon city (24.80%); 0 otherwise
 - Choongchung North = 1 if the residence area of the policyholder is Choongchung North province (3.04%); 0 otherwise
 - Choongchung South = 1 if the residence area of the policyholder is Choongchung South province or Daejeon city (8.97%); 0 otherwise
 - Jeonra North = 1 if the residence area of the policyholder is Jeonra North province (2.85%); 0 otherwise
 - Jeonra South = 1 if the residence area of the policyholder is Jeonra South province or Gwangju city (5.32%); 0 otherwise
 - Kyungsang North = 1 if the residence area of the policyholder is Kyungsang North province or Daegu city (11.82%); 0 otherwise
 - Kyungsang South = 1 if the residence area of the policyholder is Kyungsang South province or Ulsan city (13.33%); 0 otherwise
-

might be induced by the use of bonus-malus rates in lieu of the past driving history would affect both approaches in the same way.

RESULTS

Evidence of Information Asymmetry in Coverage Area

We first analyze the effects of different measurement methods while focusing on coverage areas. For the sake of comparison, we measure the coverage areas in three ways: (1) dichotomous measures with sample selection, (2) dichotomous measures without sample selection, and (3) multinomial measures. The second measurement approach has conventionally been applied in the past studies of coverage area, as in Chiappori and Salanié (2000). Nevertheless, to gauge the effect of sample selection, which can be an issue in studies of coverage amount, we start with dichotomous measures with sample selection.

For each coverage area, we construct a dichotomous coverage area variable with sample selection, which is coded 1 if a policyholder is covered by the corresponding optional coverage area and 0 if he or she is covered only by the legally mandated liability. Any policyholders who purchase optional insurance in any area(s) other than the optional coverage in question are thus excluded from the corresponding analysis. For example, those who choose one or more optional coverage areas but not collision (35.5 percent of our sample) do not appear in the sample when we examine information asymmetry in the collision coverage area.

With these dichotomous measures of coverage choices, we proceed with the two-stage estimation. The first five columns of Table 3 present the results from the second-stage accident occurrence regressions. We only report the regression coefficients and associated Murphy–Topel standard errors for the residual variables estimated from the first-stage coverage choice equation.

In all five coverage areas, the coefficients of the coverage choice residual variables fail to obtain statistical significance at the 5 percent level. If one takes this evidence obtained from the standard dichotomous measurement approach with sample selection, one would be led to conclude no information asymmetry in each coverage area when taking each coefficient in isolation.¹⁶ These outcomes are not surprising; this approach is at risk of introducing a sample selection bias because, for example, we exclude those policyholders who choose optional coverage areas other than collision from our analysis of liability versus collision (column 3 in Table 3). However, according to existing theory, these policyholders are more likely to have accidents than those with liability coverage only. So omitting them systematically underestimates the relationship between coverage and accidents, and thus results in no significant relation between them, as shown in Table 3.

¹⁶ This result does not necessarily imply the absence of information asymmetry in all coverage areas *collectively*. Albeit statistically insignificant, all five regression coefficients are positive, suggesting that there might be positive evidence for information asymmetry if all the coverage areas are considered *simultaneously*. Our multinomial measurement of insurance coverage area is designed to address this possibility and provides a simple and effective way of testing the presence of information asymmetry while considering policyholder's choices over all coverage areas simultaneously.

TABLE 3
Negative Binomial Regression Results for Asymmetric Information: Coverage Area

Residual Variables From the First-Stage Probit Regression	Dichotomous Approach							
	With Selection Bias				Without Selection Bias			
	Additional Bodily Injury	Property Damage	Collision	Personal Injury Protection	Un/Underinsured Motorist Protection	Any Optional	All Optional	Multinomial Approach
Dichotomous residual ($\hat{\epsilon}_i^b$ of Equation (2))	0.150 (0.131)	0.163 (0.130)	0.285 (0.155)	0.193 (0.131)	0.171 (0.143)	0.158 (0.130)	0.216** (0.031)	
Ordinal residual 1 ($\hat{\epsilon}_i^{m1}$ of Equation (4))								0.742* (0.348)
Ordinal residual 2 ($\hat{\epsilon}_i^{m2}$ of Equation (5))								0.182** (0.033)
Dispersion α	-0.846** (0.303)	-0.854** (0.304)	-1.438* (0.564)	-0.789** (0.292)	-0.792** (0.301)	-0.848** (0.303)	-0.489* (0.198)	-0.562** (0.208)
LR χ^2	353.9**	355.3**	367.4**	346.7**	344.4**	353.1**	442.5**	514.7**
Sample size	28,685	28,665	18,518	27,728	25,803	28,689	28,689	28,689

Notes: The Murphy–Topel estimates of the standard errors are in parentheses. For each equation, we include all the independent variables listed in Table 2, the outcomes of which are not reported here.
* and ** indicate significance at the 0.05 and 0.01 levels, respectively.

To examine the effects of sample selection, we construct a new set of dichotomous measures while including all policyholders in the sample. Specifically, we construct a dichotomous measure of coverage choices (Any Optional) that compares those who purchase any of the five optional coverage areas (coded 1) and those who purchase only the legally mandated liability (coded 0). This measure replicates Chiappori and Salanié's (2000) work. We also construct another dichotomous measure with a different cutoff point (All Optional) to compare those who purchase all optional coverage areas (coded 1) and those who do not (coded 0). Note that these measures include all policyholders in the sample.¹⁷ The sixth and seventh columns of Table 3 report the results when we use these measures.

The results differ depending on the cutoff point. The coefficient of the coverage choice residual variable remains statistically insignificant if we focus on the choice between liability only versus any optional coverage (column 6), consistent with Chiappori and Salanié (2000). However, the coefficient becomes positive and statistically significant, even at the 1 percent level, if we consider the choice between all and fewer than all optional coverage areas (column 7). Taken together, these findings suggest that theoretically predicted opportunism emerges only among those with all optional coverage areas. Those with some optional coverage are not particularly more opportunistic than those who purchase no optional coverage. More important, this finding clearly demonstrates how sample selection, whether wittingly or unwittingly, can compromise the ability to discern evidence in the data.

Although these measures of coverage choices are free from sample selection biases, they still involve excessive bundling of coverage area categories. For the any optional coverage variable in Table 3, those who purchase only some optional coverage areas get bundled together with those who purchase all areas into a single category. This approach is particularly problematic in our sample because those who purchase all optional coverage areas behave differently than others, as we show in column 7.

To examine how this excessive bundling affects the outcomes, we construct a multinomial measure of coverage area choices, which uses 0 to refer to the choice of no optional coverage area, 1 for some optional coverage areas, and 2 for all optional coverage areas. For this measure, as we discussed previously, we estimate two residuals from the first-stage coverage choice equation that enter the second-stage accident occurrence equation. The last column in Table 3 reports the results of this multinomial measure.

As in the all optional coverage measure of dichotomous coverage choice (column 7), the second residual variable (ordinal residual 2), which compares the choice of some versus all coverage areas, has a positive and statistically significant coefficient.

¹⁷ To include all policyholders in the analysis, we could measure the choice of coverage areas, coded 1 for those who choose collision coverage and 0 for all others, including those who purchase any options other than collision coverage. However, this approach violates the rank-order condition of measurements. Imagine a policyholder "A" who purchases collision coverage but no other optional coverage areas and another policyholder "B" who purchases all the optional coverage areas except collision. From the perspective of the complete coverage areas, B may be judged as covered in more areas than A is, but A is treated as having more coverage than B by this measure.

However, unlike the “any optional” coverage measure of dichotomous coverage choice (column 6), the coefficient of the first residual variable (ordinal residual 1), which compares no optional coverage areas with some optional coverage areas, is not only positive in direction but also statistically significant at the 5 percent level. Furthermore, the point estimate increases significantly from 0.158 (column 6) to 0.742 (column 8). This finding clearly demonstrates that information asymmetry is present in the choice of whether to purchase not only all optional coverage areas but also any optional coverage areas.

More important, this finding shows that a dichotomous measure, even without sample selection, introduces excessive bundling of coverage categories and thereby compromises the ability to detect the presence of information asymmetry. We cannot find information asymmetry in the choice between any versus no optional coverage areas, which shows that conventional dichotomous measurement strategy is vulnerable to some ulterior factors, such as the cutoff point used to dichotomize inherently multinomial coverage and the way cases get distributed across different levels of coverage in a particular sample. Apparently, the conventional dichotomous measurement strategy, with or without sample selection, fails to provide a robust test of information asymmetry.

Evidence of Information Asymmetry in Coverage Amount

To assess the evidence of information asymmetry in the coverage amount choices, we first replicate the conventional dichotomous measures with sample selection. Existing studies of coverage amount choices (Puelz and Snow, 1994; Cohen, 2005; Saito, 2006) exclude policyholders who choose not to buy the optional coverage area in question. We focus on a policyholder’s choice of coverage amount in following three optional coverage areas: property damage, collision, and personal injury protection.¹⁸ For each coverage area, we use the following cutoff points to separate small and large amount of coverage: property damage is coded 0 if it is covered at the minimum level (84.1 percent) and 1 otherwise; collision deductible is coded 0 if the deductible level is higher than the minimum deductible level (6.4 percent) and 1 if it is set at the minimum deductible level; and personal injury protection is coded 0 if it is covered at the minimum level (59.7 percent) and 1 otherwise.¹⁹ Despite potential arbitrariness, it must be noted that the outcomes reported in this section are robust to the use of different cutoff points.²⁰ As in previous studies of coverage amount choices, we exclude all those who opt not to purchase the coverage area in question from the corresponding analysis. Of our sample, 48.6 percent choose no collision, 16.5 percent no personal injury protection, and 13.3 percent no property damage coverage.

Following the same estimation procedure discussed in the previous section, the first three columns of Table 4 present the accident occurrence regression coefficients and

¹⁸ As we show in Table 1, virtually all policyholders choose the same level of coverage amount for additional bodily injury coverage, and no different coverage amounts are offered for uninsured motorist protection coverage.

¹⁹ We choose these cutoff points so that observations are as evenly distributed across categories as possible.

²⁰ Outcomes with different cutoffs are available upon request.

TABLE 4
Negative Binomial Regression Results for Asymmetric Information: Coverage Amount

Residual Variables From the First-Stage Probit Regression	Dichotomous Approach						Multinomial Approach		
	With Selection Bias			Without Selection Bias			Property Damage	Collision Deductibles	Personal Injury Protection
	Property Damage	Collision Deductibles	Personal Injury Protection	Property Damage	Collision Deductibles	Personal Injury Protection			
Dichotomous residual ($\hat{\varepsilon}_i^b$ from Equation (2))	-0.041 (0.033)	-0.032 (0.056)	-0.013 (0.028)	-0.009 (0.033)	0.204** (0.030)	0.034 (0.027)	0.772 (0.449)	0.128* (0.060)	0.428** (0.201)
Ordinal residual 1 ($\hat{\varepsilon}_i^{m1}$ from Equation (4))							0.003 (0.031)	0.101 (0.060)	0.000 (0.025)
Ordinal residual 2 ($\hat{\varepsilon}_i^{m2}$ from Equation (5))							-0.486** (0.197)	-0.514* (0.220)	-0.500* (0.199)
Dispersion α	-0.634** (0.220)	-1.140** (0.420)	-0.633** (0.221)	-0.369* (0.181)	-0.506* (0.220)	-0.375* (0.182)	471.7** (28.687)	433.5** (25.415)	451.3** (28.687)
LR χ^2	186.1**	103.3**	176.3**	381.6**	429.0**	383.1**			
Sample size	24,879	11,460	23,942	28,687	25,415	28,687			

Notes: The Murphy–Topel estimates of the standard errors are in parentheses. For each equation, we include all the independent variables listed in Table 2, the outcomes of which are not reported here.

* and ** indicate significance at the 0.05 and 0.01 levels, respectively.

associated Murphy–Topel standard errors for the residual variables obtained from the first-stage coverage choice regression. None of the residual variables are statistically significant at the 5 percent level. Furthermore, in all areas, the corresponding coefficients are negative. Insofar as we measure a policyholder's choice of coverage amount in a conventional dichotomous manner, we fail to find evidence of information asymmetry across different coverage areas.

However, the next three columns in Table 4 show that this conclusion is unwarranted. To eliminate the likely sample selection bias, we construct a new set of dichotomous measures that include those who opt not to purchase the corresponding coverage area (or equivalently, those who purchase a zero amount of the coverage). In particular, we code both property damage (column 4) and personal injury protection (column 5) as 0 for zero or small amounts and 1 for large amounts. To separate small and large coverage amounts, we use the same cutoff points that are used for the conventional dichotomous measures. We code collision deductibles (column 5) 0 for infinite or large deductibles and 1 for small deductibles, and we again use the same cutoff point to separate large and small deductibles.

The regression outcomes with these new dichotomous variables demonstrate that sample selection induces a bias. Although the new measures of coverage amounts fail to produce evidence of information asymmetry in property damage and personal injury protection, we observe notable changes in the regression coefficients: the magnitude of the negative coefficient decreases greatly for property damage, and the coefficient for personal injury protection turns positive. The largest impact occurs for collision. Its coefficient is not only positive, but also statistically significant at the 5 percent level, a telltale sign of information asymmetry.

Not surprisingly, the impact of sample selection bias relates closely to the number of observations incorrectly eliminated from the analysis. The extent of omission is greatest for the collision deductible (48.6 percent of the full sample removed), followed by personal injury protection (16.5 percent) and property damage (13.3 percent). The differences we observe in the findings with and without selection bias clearly reflect this varying extent of omission, and thus indicate a sample selection bias.

To assess whether the dichotomous nature of the coverage amount variables further affects our ability to discern the evidence, we finally construct a multinomial measure of coverage amount choices, such that those who choose not to purchase the corresponding coverage area are coded 0 (i.e., zero coverage amount), those who purchase a small coverage amount are coded 1, and those who purchase a large amount are coded 2. We again use the same cutoff points to distinguish between small and large coverage amounts.

As we report in the last three columns of Table 4, the results reveal that our multinomial measures of coverage amounts outperform their conventional dichotomous counterparts in terms of detecting evidence of information asymmetry. As in the analysis with dichotomous measures without sample selection, the multinomial measure of collision deductible attests to the presence of information asymmetry in this choice. Furthermore, the multinomial approach provides positive evidence of information asymmetry in the other areas of coverage. In the personal injury protection

regression, the coefficient on ordinal residual 1 is not only positive but also statistically significant at the 5 percent level. In the case of property damage, the coefficient on ordinal residual 1 turns positive, and its statistical significance improves to a p -value of 0.086. The multinomial measurement approach clearly reveals what has been obscured in dichotomous measurement approaches—that is, how much more extensive information asymmetry is in automobile insurance markets.

A comparison of outcomes between the multinomial and dichotomous approaches helps illuminate how bundling of coverage choices obscures the evidence and in what context. Note that none of the ordinal residual 2 is statistically significant in the multinomial measurement approach. This indicates that there is no significant behavioral difference between those with a small, and those with a large coverage amount in each corresponding coverage area. Although this behavioral similarity between the two groups can be attributed to the limited variation in coverage amount offered by the Korean insurance company,²¹ it can still create the problem of bundling to collapse those with small coverage amounts with those who did not purchase the coverage at all, which is the case with the dichotomous approach without sample selection. The effect of this bundling is of course compounded by the extent of bundling. As shown in Table 1, only 736 (5 percent) out of 14,734 collision policyholders opted for more than the minimum deductibles (small coverage amount). Accordingly, the extent of bundling is kept minimal here, which can explain why the conventional dichotomous measurement approach without sample selection could detect significant evidence for information asymmetry in this policy option. In contrast, the number of those policyholders who opted for a small coverage amount is substantial in other coverage areas (84 percent in property damage and 60 percent in personal injury protection). The problem of bundling is extensive here, undermining the ability to detect significant evidence of information asymmetry. It then follows that the conventional dichotomous approach may detect information asymmetry insofar as it includes no sample selection bias; however, this test strategy is sensitive to ulterior factors such as sample distribution across coverage amounts, thus failing to provide a robust test for information asymmetry.

Unlike the conventional dichotomous measure with sample selection biases, our multinomial measurement approach accurately documents the presence of information asymmetry in some of the coverage amount choices, such as collision deductibles and personal injury protection levels. In addition, the multinomial measure enables us to pinpoint more precisely which choice is subject to policyholders' opportunistic behaviors under information asymmetry. The results show that the theoretically predicted opportunistic behaviors are driven not only by the extent of coverage amount but also by whether a person is covered or not, at least in the Korean automobile insurance market.

CONCLUSION

Since Puelz and Snow (1994), increasing interest has focused on empirically testing theoretical predictions of information asymmetry, especially in automobile insurance

²¹ For instance, deductible levels (ranging from \$40 to \$476) are not so different in magnitude, compared to being insured versus being not insured.

markets in diverse countries such as the United States, Canada, France, Israel, Japan, and Taiwan. The evidence remains inconclusive with mixed outcomes, and evidence of no information asymmetry comes mostly from studies that apply the conditional correlation approach to cross-sectional data. In applying this test strategy in practice, however, researchers have reduced inherently multinomial coverage choices into dichotomous ones by focusing on whether a person purchases optional coverage or not and, once purchased, whether he or she opts for smaller or larger coverage amounts. This measurement strategy tends to cause excessive bundling and, in some situations, sample selection biases. When we replicate this approach by constructing dichotomous measures of insurance coverage for South Korean drivers, we encounter difficulty finding robust evidence of information asymmetry, either in coverage area or coverage amount choices. Positive evidence of information asymmetry emerges only when we address the sample selection bias.

Against this background, we propose an alternative approach that measures insurance coverage in an ordered multinomial scale. With this multinomial measurement approach, we find robust evidence of information asymmetry in several dimensions of insurance coverage. Furthermore, this multinomial approach enables us to describe more accurately where in the coverage choices information asymmetry occurs. In the Korean automobile insurance market, information asymmetry is prevalent in the choices of both coverage areas and coverage amounts, even after controlling for the risk classification and experience ratings used to reduce the information gap between an insured and an insurer.²² And this finding cannot be attributable to the uniqueness of the Korean automobile insurance markets because we could not detect information asymmetry when the conventional dichotomous measures of insurance coverage were used.

These outcomes clearly demonstrate the sensitivity of an empirical analysis of information asymmetry to different measurement strategies. Chiappori and Salanié's (1997, 2000) widely accepted conditional correlation approach provides a standard empirical test strategy for information asymmetry, but it too assumes that the issue of measuring insurance coverage is a nonfactor for analyzing empirical data. Our findings suggest that such an assumption is not warranted. With the multinomial measurement scheme that we devise, Chiappori and Salanié's conditional correlation approach would provide a more robust test for information asymmetry even with cross-sectional data.

APPENDIX

Murphy–Topel Variance Estimates With Multiple Predicted Values

We consider a two-stage model in which the first stage yields k predicted values to be used as covariates in the second stage. The k predicted values (we are specifically interested in the application of $k = 2$ generalized residuals) may be written as differentiable functions of the estimated linear predictor from the first stage,

²² For the theory of the effects of risk classification and experience ratings on information asymmetry, see Rubinstein and Yaari (1983) and Bond and Crocker (1991) for moral hazard, and see Dionne and Lasserre (1985), Crocker and Snow (1986), and Cooper and Hayes (1987) for adverse selection.

$\hat{\varepsilon}^{m1} = h_1(\hat{\eta}_1), \dots, \hat{\varepsilon}^{mk} = h_k(\hat{\eta}_1)$, where $\hat{\eta}_1 = X\hat{\alpha}$, and $\hat{\alpha}$ is the estimated coefficient vector in the first-stage regression.

The coefficient vector for the second stage is partitioned to emphasize the coefficients associated with the generalized residuals, $\beta = (\beta_0^T, \beta_\varepsilon^1, \dots, \beta_\varepsilon^k)^T$, where β_0 is the vector of coefficients on the observables, and $\beta_\varepsilon^j, j = 1, \dots, k$, is the scalar estimated coefficient on the generalized residuals. This partitioning helps visualize the error introduced in the second stage by the estimation of the first stage. The estimating equation for the second stage is written as $Z^T(\partial L_2 / \partial \eta_2) = 0$, where Z is a covariate matrix used in the second stage containing X and $\hat{\varepsilon}^m$, L_2 is the log likelihood of the second-stage model $\eta_2 = X\beta_0 + h_1(\hat{\eta}_1)\beta_\varepsilon^1 + \dots + h_k(\hat{\eta}_1)\beta_\varepsilon^k$, and Z is partitioned on the same columns on which we partitioned β .

Because the second-stage model estimation includes predicted values from the estimation of the first stage, we must take into account the initial estimation when generating an estimated variance matrix for the model coefficients in the second stage. The Murphy–Topel two-stage, model-based variance estimator is defined as $V_{MT} = V_2 + V_2 [DV_1 D^T - DV_1 R^T - RV_1 D^T] V_2$, where V_i is the model-based variance estimator for the i th stage, $D = E[(\partial L_2 / \partial \beta)(\partial L_2 / \partial \alpha^T)]$ and $R = E[(\partial L_2 / \partial \beta)(\partial L_1 / \partial \alpha^T)]$, where L_i is the log likelihood of i th stage model.

A general derivation of the components of this variance estimator in terms of generalized residuals as differentiable functions of the linear predictor is given by

$$\begin{aligned} \hat{D} &= \sum_{i=1}^n \left(\frac{\partial L_2}{\partial \beta} \right)_i \left(\frac{\partial L_2}{\partial \alpha^T} \right)_i = \sum_{i=1}^n \left(\frac{\partial L_2}{\partial \beta} \right)_i \left(\frac{\partial L_2}{\partial \eta_2^T} \frac{\partial \eta_2^T}{\partial \eta_1^T} \frac{\partial \eta_1^T}{\partial \alpha^T} \right)_i \\ &= Z^T \left[\text{Diag} \left(\frac{\partial L_2}{\partial \eta_2} \frac{\partial L_2}{\partial \eta_2^T} \sum_{j=1}^k \hat{\beta}_\varepsilon^j \frac{\partial h_j(\hat{\eta}_1^T)}{\partial \hat{\eta}_1^T} \right) \right] X \end{aligned}$$

and

$$\begin{aligned} \hat{R} &= \sum_{i=1}^n \left(\frac{\partial L_2}{\partial \beta} \right)_i \left(\frac{\partial L_1}{\partial \alpha^T} \right)_i = \sum_{i=1}^n \left(Z^T \frac{\partial L_2}{\partial \eta_2} \right)_i \left(\frac{\partial L_1}{\partial \eta_1^T} X \right)_i \\ &= Z^T \left[\text{Diag} \left(\frac{\partial L_2}{\partial \eta_2} \frac{\partial L_1}{\partial \eta_1^T} \right) \right] X. \end{aligned}$$

Armed with the model-based variance estimators for each of the two model stages, along with estimates of the D and R matrices, the Murphy–Topel variance estimator is easy to compute.

REFERENCES

- Abbring, J., P. A. Chiappori, and J. Pinquet, 2003, Moral Hazard and Dynamic Insurance Data, *Journal of European Economic Association*, 1: 767-820.
- Akerlof, G., 1970, The Market for Lemons: Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*, 84: 488-500.

-
- Arnott, R., and J. Stiglitz, 1988, The Basic Analytics of Moral Hazard, *Scandinavian Journal of Economics*, 90: 383-413.
- Bond, E., and K. Crocker, 1991, Smoking, Skydiving and Knitting: The Endogenous Categorization of Risks in Insurance Markets With Asymmetric Information, *Journal of Political Economy*, 99: 177-200.
- Boyer, M., and G. Dionne, 1989, An Empirical Analysis of Moral Hazard and Experience Rating, *Review of Economics and Statistics*, 71: 128-134.
- Cardon, J., and I. Hendel, 2001, Asymmetric Information in Health Insurance: Evidence From the National Medical Expenditure Survey, *RAND Journal of Economics*, 32, 408-427.
- Chiappori, P. A., 2000, Econometric Models of Insurance Under Asymmetric Information, in: G. Dionne, ed., *Handbook of Insurance* (Boston, MA: Kluwer Academic), pp. 365-394.
- Chiappori, P. A., F. Durand, and P. Y. Geoffard, 1998, Moral Hazard and the Demand for Physician Services: First Lessons From a French Natural Experiment, *European Economic Review*, 42: 499-511.
- Chiappori, P. A., B. Jullien, B. Salanié, and F. Salanié, 2006, Asymmetric Information in Insurance: General Testable Implications, *RAND Journal of Economics*, 37: 783-798.
- Chiappori, P. A., and B. Salanié, 1997, Empirical Contract Theory: The Case of Insurance Data, *European Economic Review*, 41: 943-950.
- Chiappori, P. A., and B. Salanié, 2000, Testing for Asymmetric Information in Insurance markets, *Journal of Political Economy*, 108: 56-78.
- Cohen, A., 2005, Asymmetric Information and Learning: Evidence From the Automobile Insurance Market, *Review of Economics and Statistics*, 87: 197-207.
- Cooper, R., and B. Hayes, 1987, Multi-Period Insurance Contracts, *International Journal of Industrial Organization*, 5: 211-231.
- Crocker, K., and A. Snow, 1986, The Efficiency Effects of Categorical Discrimination in the Insurance Industry, *Journal of Political Economy*, 94: 321-344.
- Cummins, J. D., and S. Tennyson, 1996, Moral Hazard in Insurance Claiming: Evidence From Automobile Insurance, *Journal of Risk and Uncertainty*, 12: 29-50.
- De Meza, D., and D. Webb, 2001, Advantageous Selection in Insurance Markets, *RAND Journal of Economics*, 32: 249-262.
- Dionne, G., and R. Gagné, 2002, Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance, *Journal of Risk and Uncertainty*, 24: 213-230.
- Dionne, G., C. Gouriéroux, and C. Vanasse, 2001, Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment, *Journal of Political Economy*, 109: 444-453.
- Dionne, G., and P. Lasserre, 1985, Adverse Selection, Repeated Insurance Contracts and Announcement Strategy, *Review of Economic Studies*, 52: 719-723.
- Dionne, G., P. C. Michaud, and M. Dahchour, 2006, Separating Moral Hazard From Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence From France, Working Paper, SSRN.
- Dionne, G., and C. Vanasse, 1992, Automobile Insurance Ratemaking in the Presence of Asymmetrical Information, *Journal of Applied Econometrics*, 7: 149-165.

- Finkelstein, A., and J. Poterba, 2004, Adverse Selection in Insurance Markets: Policyholder Evidence From the U.K. Annuity Market, *Journal of Political Economy*, 112: 183-208.
- Hardin, J., 2002, The Robust Variance Estimator for Two-Stage Models, *Stata Journal*, 2: 253-266.
- Hendel, I., and A. Lizzeri, 2003, The Role of Commitment in Dynamic Contracts: Evidence From Life Insurance, *Quarterly Journal of Economics*, 118: 299-327.
- Holly, A., L. Gardiol, G. Domenighetti, and B. Bisig, 1998, An Econometric Model of Health Care Utilization and Health Insurance in Switzerland, *European Economic Review*, 42: 513-522.
- Jullien, B., B. Salanié, and F. Salanié, 2007, Screening Risk-Averse Agents Under Moral Hazard: Single-Crossing and the CARA Case, *Economic Theory*, 32: 151-169.
- Li, C., C. Liu, and J. Yeh, 2007, The Incentive Effects of Increasing Per-Claim Deductible Contracts in Automobile Insurance, *Journal of Risk and Insurance*, 74: 441-459.
- Murphy, K. M., and R. H. Topel, 1985, Estimation and Inference in Two-Step Econometric Models, *Journal of Business & Economic Statistics*, 3: 370-379.
- Puelz, R., and A. Snow, 1994, Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market, *Journal of Political Economy*, 102: 236-257.
- Richaudeau, D., 1999, Automobile Insurance Contracts and Risk of Accident: An Empirical Test Using French Individual Data, *Geneva Papers on Risk and Insurance Theory*, 24: 97-114.
- Rothschild, M., and J. Stiglitz, 1976, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics*, 90: 629-649.
- Rubinstein, A., and M. Yaari, 1983, Repeated Insurance Contracts and Moral Hazard, *Journal of Economic Theory*, 30: 74-97.
- Saito, K., 2006, Testing for Asymmetric Information in the Automobile Insurance Market Under Rate Regulation, *Journal of Risk and Insurance*, 73: 335-356.
- Vera-Hernandez, M., 2003, Structural Estimation of a Principal-Agent Model: Moral Hazard in Medical Insurance, *RAND Journal of Economics*, 34: 670-693.
- Wang, J. L., C. Chung, and L. Y. Tzeng, 2008, An Empirical Analysis of the Effects of Increasing Deductibles on Moral Hazard, *Journal of Risk and Insurance*, 75: 551-566.

Copyright of Journal of Risk & Insurance is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.