Two Tests for ex ante Moral Hazard in a Market for Automobile Insurance

Author

Rowell, David, Nghiem, Son, Connelly, Luke B

Published

2017

Journal Title

Journal of Risk and Insurance

Version

Accepted Manuscript (AM)

DOI

10.1111/jori.12161

Two Tests for *ex ante* Moral Hazard in a Market for Automobile Insurance

David Rowell[1], Hong Son Nghiem[2] & Luke B Connelly[3]

The University of Queensland, Queensland Brain Institute, Asia Pacific Centre for Neuromodulation[1]
Queensland University of Technology, Institute of Health and Biomedical Innovation[2]
Faculty of Health and Behavioural Sciences and Centre for the Business and Economics of Health, The University of Queensland[3]

Abstract

Empirically separating the phenomena of moral hazard and adverse selection in insurance markets has occupied researchers in this field for decades. Recently the potential benefits of using survey data instead of claims data to control for the different dimensions of private information when testing for evidence of asymmetric information have been explored in the insurance literature. This paper extends that approach to present two tests for *ex ante* moral hazard in a market for automobile insurance. In this paper we specify (i) a recursive model and (ii) an instrumental variables model to address endogeneity with respect to policy selection in cross-sectional road traffic crash (RTC) survey data. We report a statistically significant *ex ante* moral hazard effect with both models. This result is then subjected to a falsification test whereby the analysis is repeated in sub-samples of at-fault and non-at-fault RTCs. Our anti-test produces no evidence of *ex ante* moral hazard in the sub-sample of not-at-fault RTCs, in which the true moral hazard may reasonably be assumed to be zero, thus supportingthe interpretation of the results of our two models. Our extension of the existing literature via these two specifications may have useful analogs in other insurance markets for which survey data are available.

Acknowledgment: First Author acknowledges financial support the Brian Gray scholarship program, which is jointly funded by the Australian Prudential Regulation Authority (APRA) and the Reserve Bank of Australia (RBA).

Introduction

Moral hazard arises when the probability or size of a loss increases due to the insured's subsequent behaviour (Hölmstrom 1979). Winter (1992) stated that two conditions must be satisfied for *ex ante* moral hazard to occur in a market for insurance. First, the probability that the insured event occurs can be influenced by the level of preventive effort expended after the contract has been signed. Second, the level of preventive effort cannot be monitored and included in the contract. Arrow (1985) identifies two types of principal-agent problems in markets for insurance; hidden action and hidden information. Hidden action (moral hazard) occurs because preventive effort, which is a disutility to the agent, affects the probability of an insured event. Hence insurance provides an incentive for the agent to reduce preventive effort, thereby generating a positive correlation between insurance and claim. Hidden information (adverse selection) occurs when the agent has some private knowledge about the probability of the insured event, which they use to inform their decision to purchase insurance. The differentiation of moral hazard from adverse selection is empirically challenging because both phenomena are associated with a positive correlation between the decision to insure and the probability of a claim. Indeed, Cohen and Siegelman (2010, p. .71) have characterised "[t]he disentanglement of adverse selection and moral hazard is probably the most significant and difficult challenge that empirical work on adverse selection [or moral hazard] in insurance markets faces."

This paper attempts to disentangle moral hazard from adverse selection in a market for road traffic crash (RTC) insurance and invokes two strategies to do so. First, we identify a survey dataset that includes both claimants and non-claimants. Critically, this dataset contains data fields that are not within the insurer's information set. While the existing literature has relied on claims data, we employ household data that was originally collected for market research on the Australian smash repair market. Importantly, this dataset includes information on the respondent's insurance policy (if the driver is insured), driver demographics, vehicle characteristics, RTC incidence and history. Second, we use two econometric strategies that are commonly used to address a simultaneity problem and have not previously been applied in this literature. We specify a recursive model and a bivariate probit model with an instrumental variable to address the likely endogeneity of the policy type with respect to the probability distribution of the loss. In the latter approach, the household's insurance decision with respect to a secondary vehicle at the household is used to instrument for the insurance decision on the primary vehicle. Briefly, the rationale for this choice of instrument (which we explain in greater detail below) is that, while the choices of insurance type on both vehicles is correlated, no amount of insurance on the secondary vehicle may reasonably be expected to affect the insured's behaviour when driving the primary vehicle.

We find evidence of *ex ante* moral hazard in the Australian market for vehicle insurance and subject our findings to a sensitivity analysis, which tests for moral hazard in sub-samples of motorists involved in at-fault, and not-at-fault RTCs. The basis for the choice of this sensitivity test is that the theory of moral hazard predicts a stronger correlation between insurance and at-fault RTCs than not-at-fault RTCs. The results of our sensitivity analysis are consistent with this hypothesis.

Current Evidence

The modern debate on asymmetric information in auto insurance markets can be traced to the work of Puelz and Snow (1994), which used individual claims data to construct an ordered logit model and revealed a correlation, conditional upon the insurer's information set, between risk-type and choice of deductible. Puelz and Snow (1994) argued that the statistically significant negative sign of this correlation coefficient constitutes empirical evidence of adverse selection. It has been argued, though, that this result is also consistent with the hypothesis of *ex ante* moral hazard (Chiappori, 1999) and that the model was incorrectly specified (Dionne et al., 2001), which could lead to the identification of a conditional correlation when there is none. Dionne et al. (2001) demonstrate that when the nonlinearity of the risk classification variables are accounted for, the residual correlation, which was interpreted as adverse selection, vanishes.

Chiappori and Salanié (2000, p. 66) proposed an alternative test for asymmetric information using a bivariate probit model wherein the first probit equation predicts the level of insurance and the second probit equation predicts the occurrence of a claim. The null hypothesis of no asymmetric information was tested using two null hypotheses: (i) zero covariance of residuals when estimating two probit estimators separately; and (ii) zero correlation coefficient of the two residuals when the model is estimated as a bivariate probit. Using a French claims dataset that contains 55 dummy variables, to control for the insurer's information set Chiappori and Salanié (2000) report no evidence of asymmetric information in this sub-population of beginner drivers.[1] Chiappori and Salanié (2000) conclude with a specific test for moral hazard that exploits a 'natural experiment' whereby adult children can inherit their parent's *bonus-malus* coefficient if they state that their automobile is jointly owned. A dichotomous *bonus-malus* variable equal to one if the beginner driver inherits a *bonus-malus* coefficient of 0.5 is added to the coverage and claims equations. Chiappori and Salanié (2000) argue that the sign of the coefficient on the *bonus-malus* variable from the claims equation may be used to identify the presence *ex ante* moral hazard. They report a negative coefficient, which leads them to accept the hypothesis of no moral hazard in that dataset.

[1] Subsequent analysis of German claims data by Su (2013) and Dutch claims data by Zavadil (2014) found no evidence of asymmetric information. However, Spindler (2014) did report asymmetric information in a sub-set of German policyholders

Following this study two distinct methodological responses can be identified. Firstly, Dionne et al. (2013) argued that by limiting the analysis to beginner drivers, Chiappori and Salanié (2000) had omitted a measure of claims history that may conceal a conditional correlation, because this variable is both negatively correlated with contract choice and positively correlated with claims. Rather than analyzing claims data, Dionne et al. (2013) used a three-year panel of survey data collected by the French market research firm SOFRES, to test for *ex ante* moral hazard using the conditional correlation approach. They specified a bivariate probit model where the probabilities of accidents and insurance contract choice in the current period are the function of both outcomes in the previous period and other covariates. To specify a test for *ex ante* moral hazard they Dionne et al. (2013) argued:

> [w]e assume drivers differ in terms of risk type (or ability). In the model, agents first buy insurance without knowledge of their own risk. They learn about risk from their history of accidents. Accidents differ depending on whether the driver is at fault or not. Although the insurer observes the *bonus-malus* he does not learn as fast as the agent about his riskiness, partly because he/she observes claims only. Thus asymmetric learning develops, which may lead to pure adverse selection in contract choices within observable risk classes (Dionne et al., 2013, p. 900)

A Granger causality test was used to test for *ex ante* moral hazard by examining the correlation between previous contract choice and an accident in the current period, conditional upon the insurer's information set. The authors report strong evidence of *ex ante* moral hazard for drivers with less than 15 years of driving experience. Importantly the ability to include an indicator of historical RTCs that were not observable by the insurer enabled Dionne et al. (2013) to specify their test for *ex ante* moral hazard.

A second approach promoted by Abbring et al. (2003) and Israel (2004) eschewed testing for conditional correlation using cross-sectional data, instead choosing to analyse longitudinal data. These authors argued that while the conditional correlation approach offers a robust test for asymmetric information it cannot be used to distinguish moral hazard from adverse selection. Abbring et al. (2003) adapted the state-dependence approach by Heckman and Borjas (1980) to test for moral hazard. A proportional hazard model was used to compare: (i) the distribution of first and second claims across contracts over time, and (ii) the first and second claim times of each contract with two claims (or more) in a French claims dataset. No evidence of moral hazard was reported. Israel (2004), however, has argued that Abbring et al. (2003) assumed that there are no other sources of state dependence. In particular, past accidents were explicitly assumed only to influence current behavior through their effects on the premium. Israel (2004) therefore used a difference-in-difference approach to examine the occurrence of claims around the three-year insurance event and report a small but statistically significant *ex ante* moral hazard effect.[2]

Weisburd (2015) conducted tests for *ex ante* moral hazard, which exploited an Israeli dataset whereby automobile insurance coverage was distributed as an employee benefit, to 67 per cent of motorists in their sample. It was argued that the allocation of company coverage, provided an exogenous variation in insurance coverage that was independent of adverse selection. Employer-determined coverage was estimated to reduce the average cost of an accident by $235. After controlling for driver and vehicle characteristics, Weisburd (2015) reported a $100 reduction in accident costs resulted in a 1.7 percentage point increase in the probability of an accident. Given an average accident rate of 16.3 percent, it was estimated that *ex ante* moral hazard had caused a 10 percent increase in automobile accidents. A sensitivity analysis found no moral hazard with large (road) accidents vis-à-vis small (parking) accidents. It was argued that as the

[2]We direct interested readers to the recently published survey article on testing for adverse selection in markets for automobile insurance written by Cohen and Siegelman (2010)

principal cost of a parking accident was financial rather than involve personal injury, moral hazard was more likely to be observed in small accidents.

While there has been considerable debate regarding the ability of various approaches to disentangle adverse selection from moral hazard in the automobile insurance literature, the current consensus in this literature is that tests for *ex ante* moral hazard can be problematic when cross-sectional data are used (Chiappori and Salanié, 2000; Dionne et al., 2013; Abbring, 2003). However, the analysis of cross-sectional survey data does offer some promise in this field for the reason that some fields collected for surveys may reveal useful information about individuals that is typically not visible to insurers. Finkelstein and McGarry (2006) have demonstrated this point using data on the market for long-term care insurance and their work has some important implications for the identification of *ex ante* moral hazard more generally. Finkelstein and McGarry (2006) argued that the Health and Retirement Survey not only provides a rich description of the insurer's information set, but that it also includes other variables that, although germane to insurance and insured choices, are not typically observed by insurers. Finkelstein and McGarry (2006) first demonstrated that zero evidence of a conditional correlation between insurance and accident/claim (a traditional test asymmetric information) is not a necessary condition for the existence of the other dimensions of private information. Secondly, they demonstrated that the inclusion of variables that may proxy for private information that is not normally observable to the insurance firm, could provide useful insights into behavior under insurance.

In this paper, we also use survey data rather than claims data for the same basic reasons that were advanced by Finkelstein and McGarry (2006). We test for moral hazard in the Australian market for comprehensive insurance using cross-sectional data to disentangle moral hazard from adverse selection by using a recursive model and a bivariate probit model with a novel choice of instrumental variable.

Data

Over a six-week period commencing in October 1999, EKAS Marketing Research Services conducted market research on behalf of IMRAS Consulting to analyze community attitudes to the Australian smash repair market. The resulting data, henceforth referred to as the IMRAS dataset, used computer assisted telephone interviews to contact 37,833 rural and metropolitan households in four Australian states (NSW, Victoria, Queensland and WA). Vehicle owners from 4,005 households (16.9 percent) completed the survey. These data were commercially available, and purchased for this study.

Although these data were not collected for the purpose of conducting research on insurance markets *per se*, many of the variables that are necessary to analyze asymmetric information, moral hazard and adverse selection are available. Indeed, this dataset contains data on several variables that are particularly useful to the purpose at hand. Firstly, the survey included a question about the historical involvement of the respondent in any RTC. A two-year recall period was used for this question both to ensure that sufficient data were collected on RTCs and smash repair experiences, and to minimise errors in respondent recall. In total, 994 of the respondents (24.8 percent) stated that they were involved in at least one RTC during the previous two-year period. We use this field of the dataset to create the dichotomous variable RTC, equal to unity if any RTC was reported between 1997 and 1999.

Secondly, the IMRAS survey collected data on the insurance status of the respondent's automobile, as either (i) compulsory third-party (personal injury) only, (ii) (i) plus third-party property, (iii) (i) plus third-party property, fire and theft or (iv) (i) plus comprehensive property insurance. Only comprehensive insurance, which is discretionary, indemnifies the owner for the cost of smash repairs in a RTC for which he/she is at fault. Australia's comprehensive insurance policy is analogous to the French *assurance tous risques* as described by Chiappori and Salanié (2000). Importantly the dataset also enabled the identification of 20 respondents who were uninsured when their RTC occurred and subsequently purchased comprehensive insurance. A dichotomous variable *INS* was created using this field and is equal to one if the respondent indicated that they were comprehensively insured and equal to zero if they were not comprehensively insured.

Thirdly, to conduct reliable tests for *ex ante* moral hazard using conditional correlation, it is necessary to define a set of covariates that accurately reflects the insurers' information set. Two sources of information were reviewed. The first was the empirical literature, which identifies (as covariates) the data commonly collected by predominantly French insurers on their policyholders. Secondly, data collected by Australian insurance industry was reviewed. The five most frequent insurance carriers for survey respondents were the NRMA Ltd., AAMI, GIO, RACV and Suncorp. These firms, which provided cover for 58.7 percent of the sample, each hosts a web page that enables the user to obtain a quote for comprehensive insurance. There is considerable congruence between the categories of data that are recognized as important in the (i) empirical and theoretical literature; (ii) data collected by insurance firms to generate premium quotations; and (iii) data included in the IMRAS dataset. Controls for driver characteristics in the data included age, gender, young co-driver ($<$ 25 years of age), vehicle ownership type (i.e., private or corporate), location (metropolitan or rural), socioeconomic status (SES)[3] and years licensed, while controls for vehicle characteristics included the vehicle's value, age, make, body-type and engine size (4-, 6- and 8-cylinder). The literature has also emphasized the importance of including a measure of claims history. We used a dichotomous variable $RTC_{1994-97}$ (which equals one if an RTC occurred between October 1994 and October 1997 and zero otherwise in preference to the no-claim bonus variable because it is applicable both to insured and uninsured drivers. The use of this variable also obviates any concerns about differences in the insurance rules that insurers may apply to awarding no-claim bonuses.

While we believe that this set of covariates provides a good approximation of the insurer's information set, it is possible that Australian insurers also collect and use data fields that are unavailable within the IMRAS dataset to risk-rate their policyholders. For example, insurers are observed to collect data identifying whether or not the vehicle was garaged and the billing period i.e., yearly or six-monthly. These data may be used by insurers to identify risk-types more accurately. Finkelstein and McGarry (2006) have stated that a conservative approach should be adopted when selecting covariates. In their analysis of the long-term care insurance market, they included variables that are not necessarily collected by all insurers, to ensure that their model adequately accounted for the insurer's information set. Analogously we identified two variables within the IMRAS dataset: income and occupation type, which were not necessar-

[3]A measure of SES was obtained for each postcode from the Socioeconomic Indices for Areas (SEIFAs) developed by the Australian Bureau of Statistics (ABS)(2006). The ABS reports SEIFA indices by collection district (CD), which are the geographical regions used to gather census data. Each postcode is comprised of a number of collection districts. A weighted SEIFA index of Advantage-Disadvantage ($P.C._{AD\_index}$) was constructed for each postcode (PC) as follows

$$PC_{AD-index} = \frac{\sum(\frac{SEIFA_{Pop.} \times SEIFA_{AD-index}}{Pop.})}{\frac{CD}{PC}}$$

The term in parentheses is a weighted SEIFA Advantage-Disadvantage index for each postcode that controls for the estimated resident population. A categorical variable, which measures the latent SES, was constructed using quartiles of the constructed index.

ily collected by all Australian insurers, but were identified as potential proxies for data otherwise collected by insurers to risk-rate their policyholders. As it has been established that tests for asymmetric information that use conditional correlation may produce spurious results if the explanatory covariates are inappropriately specified as a linear function of RTC, we have used combinations of dummy variables to reflect risk classification and, where data are continuous, flexible approximations (e.g. spline functions) have been substituted as recommended by Dionne et al. (2013).

Recall that the advantage of survey data, in this context, is that the universe of fields available is greater than that typically collected by insurers. Table 1 presents a comparison of the available data in our survey dataset with a generic claims dataset. Dionne et al. (2013) has argued that the ability to analyze all RTCs and not just claims against RTCs, was the key, in that study, to devising a credible test for *ex ante* moral hazard. The comparison provided in Table 1 shows that our survey dataset also presents us with this possibility for the current study.

Table 1: Observable Information in Survey and Claims Data

Survey Data (IMRAS)

		RTC		Claim		**X**	**Y**
		Yes	No	Yes	No		
Insured	Yes	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
	No	$\surd$	$\surd$	n.a.	n.a.	$\surd$	$\surd$

A Generic Claims dataset

		RTC		Claim		**X**	**Y**
		Yes	No	Yes	No		
Insured	Yes	$\times$	$\times$	$\surd$	$\surd$	$\surd$	$\times$
	No	$\times$	$\times$	n.a	n.a.	$\times$	$\times$

Note: $\surd$= Data observed; $\times$ = Data unobserved; **X** = Insurer's information set; **Y** = Data unobserved by the insurer

Econometric Approach

Theoretical Model

Arrow (1985) states that the relationship between insurance and an insurable event (e.g. a RTC) can be confounded by two principal-agent problems. The first, adverse selection, occurs when unobserved risk type RT^* affects the agent's decision to purchase insurance (INS) and the subsequent probability of a RTC. The second, *ex ante* moral hazard, occurs because the agent's unobserved preventive effort PE^*, is affected by the decision to insure, which affects the subsequent probability of a RTC. Importantly, in Arrow's (1985) description of a market for insurance, the adverse selection effect precedes the moral hazard effect. The dynamics of two principal-agent problems described by Arrow (1985) can be understood as the result of multiple characteristics and a single consequence. Multiple unobserved driver and vehicle characteristics coalesce to create the hidden information problem, which may jointly affect both the insurance decision and the probability distribution of the insured event losses. However, purchasing insurance has a single consequence that manifests as reduced preventive effort, which may increase the probability of a RTC.

While Arrow (1985) does not present a formal theoretical model of a market for insurance as such, the temporal association between hidden action and hidden information described in his discussion of asymmetric information was used to model the relationship between insurance and RTCs using a path diagram (see Figure 1). . As per Pearl (2009), the solid lines denote causal links and arrows causal direction. The dashed arcs identify confounding processes. Unobserved risk type affects the insurance decision, which in turn can affect unobserved preventive effort, and both can affect the probability distribution of losses due to the insured event.

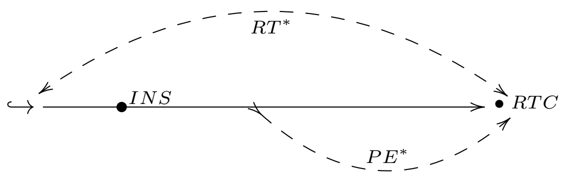


Figure 1: A path diagram for a market of insurance

The relationships illustrated in Figure 1 may be written as follows:

$$RTC_{it}^{*} = \alpha_0 + \alpha_1 INS_{it} + \alpha_2 \mathbf{X}_{it} + RT_i^{*} + PE_i^{*} + \varepsilon_{it} \tag{1}$$

where RTC_{it}^{*} is the probability of a crash of individual i during the period October 1997 to October 1999 henceforth denoted by the subscript t; and INS_{it} is an observed dummy variable, equal to one if the individual had purchases comprehensive insurance. The relationship between the latent and observed variables is that RTC_{it}^{*}>0 then RTC_{it}=1, otherwise RTC_{it}=0; INS_{it}>0 then INS_{it}=1, otherwise INS_{it}=0. $\mathbf{X}_{it}$ is a set of exogenous variables, which capture all risk factors known to the insurer, $RT_i^{*} + PE_{it}^{*} + \varepsilon_{it}$ is composite error term, which consists of unobserved risk type RT_{it}^{*}, unobserved preventive effort PE_{it}^{*} and a random noise term ε_{it}. Two remarks can be made about equation (1). First, the presence of RT_{it}^{*} and PE_{it}^{*} suggests that the error term is correlated with both INS_{it} and RTC_{it} and therefore INS_{it} is endogenous. Hence, if estimated, the coefficient α_1 would provide a biased estimate of *ex ante* moral hazard. Secondly, if RT^{*} (Arrow, 1985) and PE^{*} (Winter, 1992) were perfectly observable, then $\alpha_1$will equal zero. Insurance *per se* has no effect on the probability of a RTC.

As Dionne (2013) has argued, when the task is to estimate an empirical model of insurance that contains a potentially-endogenous variable,

[i]t is often better to instrument this variable (see Dionne et al. 2009, 2010, and Rowell 2011, for more details).

We consider two methods, previously described by Rowell (2011),[4] to obtain a consistent estimate of α_1 from equation (1). The first method, uses a recursive system model with the pre-determined variable, RTCs in the previous period, to control for unobserved risk type. Controlling for RT^* in equation (1) would enable moral hazard to be identified α_1. The second method uses the insurance status of the secondary vehicle in two-vehicle households as an instrumental variable. Theoretically, the extent to which (i) or (ii) constitutes a better specification of the model depends on the extent to which one believes that the insurer's information set is likely to be exhaustive with respect to the classification of risk types. The models are described in detail as follows.

Recursive Model

Past RTCs are correlated with the incidence of RTCs in the current period. Recall that, in our dataset, historical RTCsthat occurred between October 1994 and October 1997 are identified. Henceforth, these will be denoted with the notation $RTC_{i,t-1}$. It can be argued that the variable $RTC_{i,t-1}$ has two important properties that are necessary for its use in a recursive model. First, it proxies unobserved risk type, and hence is able to capture the adverse selection effect. Second, it is predetermined in the sense that it proxies the lag of RTC, and hence is exogenous by definition. Dionne et al. (2013) argued that RTCs not resulting in a claim enabled the insured motorist to acquire private information about their risk type. We also use past RTCs (i.e., $RTC_{i,t-1}$) as a proxy for

[4]"Moral Hazard: Empirical Evidence in the Australian Market for Automobile Insurance" (Rowell 2011) was a thesis accepted for a degree of Doctor of Philosophy at The University of Queensland

unobserved risk-type to specify a recursive model, as follows.

$$RTC_{it}^{*} = \alpha_0 + \alpha_1 INS_{it} + \alpha_2 \mathbf{X}_{1it} + RT_{it}^{*} + PE_{it}^{*} + \varepsilon_{it} \tag{2}$$

$$INS_{it}^{*} = \beta_0 + \beta_1 RTC_{i,t-1} + \beta_2 \mathbf{X}_{2it} + RT_{it}^{*} + \mu_{it} \tag{3}$$

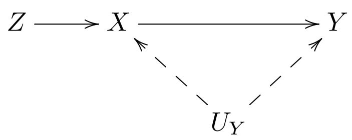
In this system, the null hypothesis of no moral hazard is given by $H_0 : \alpha_1 = 0$ in equation (2). The coefficient β_1 in equation (3) captures the adverse selection effect (i.e., high risk drivers are more likely to buy comprehensive insurance). Greene (2008, p. 407) showed that the system of equations above can be estimated efficiently and consistently using bivariate probit with full information maximum likelihood (FILM). Note that this system equation can be estimated with no exclusions (i.e., X_{1it} and X_{2it} are identical). The extent to which the parameter β_1 in equation (3) captures adverse selection is determined by the degree to which $RTC_{i,t-1}$can capture unobserved risk type, RT_{it}^{*}.

The capacity of past RTCs to capture unobserved risk type within our dataset depends on the type of RTCs reported. Conceptually, the variable $RTC_{i,t-1}$ is comprised of RTCs of two types (i) those which were reported to the insurance firms via claims and (ii) those for which no claim was made. If the variable $RTC_{i,t-1}$. includes a comprehensive array of minor RTCs, which are otherwise unobserved by insurance firms, then unobserved risk-type RT_{it}^{*} will, in part, be captured. Alternatively, if $RTC_{i,t-1}$is fully observable to the insurers then no new information identifying risk-type is provided and hence β_1 does not reflect adverse selection effect. The proportion of RTCs occurring three to five years ago that are observable to insurers generally is unknown although, anecdotally, insurers commonly ask applicants for policies to report their claims over the past three years. However, in the current period, 65.2 per cent of insured drivers who reported an RTC also reported lodging an insurance claim. Provided reporting behavior is constant over time, the variable $RTC_{1994-97}$ does provide some additional information on risk-type. Alternatively, one could argue that recall bias ensures that only major RTCs are recalled and no new information on unobserved risk-type is provided. To account for the latter possibility, we propose an alternative econometric approach, which is to estimate moral hazard while controlling for adverse selection using an instrumental variable approach.

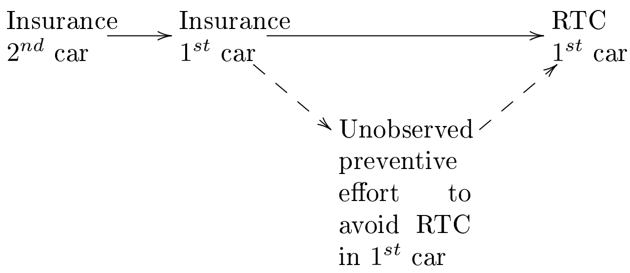
Instrumental Variable Model

Equation (1) is a probit model where RTC_{it} is a function of INS_{it}, a variable, which is binary and endogenous. A model with an endogenous variable of this type can be estimated as a bivariate probit with full information maximum likelihood (FILM)(Greene, 2008, p. 407). Note that the bivariate probit model specified above, like those used elsewhere in this literature (e.g., Chiappori and Salanié 2000; Cohen 2005; Dionne et al. 2013) has no exclusion restrictions. To ensure that our model is just identified, one instrumental variable is required. To be a credible IV, the candidate variable must not be observed and collected by the insurer, otherwise one would expect the insurer to use the observable information to rate the premium, although it is known that exceptions exist (Finkelstein and McGarry 2006). An analysis of claims data, supplied by an insurance firm, would preclude the identification of an effective IV. Our household survey dataset, though, contains information that is not typically observed by insurers, and we exploit that fact to identify an appropriate IV strategy.

To be valid an IV should satisfy two criteria; (i) it should be correlated with the endogenous variable and (ii) it should be uncorrelated with the error term (Wooldridge, 2000). Figure 2 presents a path diagram from Pearl (2009) to outline the requirements that a legitimate instrumental variable should satisfy.



(a) Generic IV (Pearl 2009 p. 248)



(b) IV for insurance

Figure 2: Path diagram for an instrumental variable

In Figure 2a, Y is the dependent variable and X an explanatory variable, which is endogenous because the effect that X has on Y is confounded by the unobserved effect U_Y has on both X and Y. More formally:

> [t]he traditional definition qualifies a variable Z as instrumental (relative to the pair (X, Y)) if (i) Z is not independent of X and (ii) Z is independent of all variables (including the error term) that have an influence on Y that is not mediated by X. Pearl (2009, p.247)

Figure 2b applies the relationships illustrated in 2a to the setting of a market for comprehensive automobile insurance. The dependent variable Y corresponds to the probability of an RTC in the primary vehicle (RTC 1^{st} car). The explanatory variable X corresponds to the probability that the primary vehicle was comprehensively insured (Insurance 1^{st} car). The economic theory states that *ex ante* moral hazard occurs when insurance firm is unable to observe or contract for the agent's preventive effort (Hölmstrom, 1979; Marshall, 1976; Mirrlees, 1999; Pauly, 1974; Shavell, 1979; Winter, 1992). Thus, U_Y corresponds to the unobserved preventative effort used to avoid a RTC in the primary vehicle (Unobserved preventative effort to avoid RTC in 1^{st} car).

Figure 2b indicates that the selected instrument should be correlated with the decision to insure the primary vehicle but not affected by the level of preventive effort invested to avoid a RTC in the primary car. We argue that the insurance status of the secondary vehicle garaged in the household satisfies both criteria and as such is a legitimate instrument for the insurance status of the primary vehicle. Firstly, the insurance status of the secondary vehicle is correlated with insurance status of the primary vehicle, which we establish empirically (further details follow). Secondly, the insurance status of the secondary vehicle is not correlated with the preventative effort expended to avoid a RTC in the primary vehicle. Our argument is presented in detail, below.

The first condition, that the insurance status of the vehicles garaged within the same household could be correlated, is satisfied for at least two reasons. Firstly, driving abilities may be familially-correlated and therefore within-household decisions to purchase comprehensive insurance. Chiappori and Salanié (2000) provided empirical evidence of a familial relationship in respect of driving abilities. Secondly, in a household where driving abilities are not correlated but use of the vehicles is shared, a correlation between choices of insurance is likely to develop. Typically, vehicle owners share access to their vehicle with their spouse, adult child or other household members. Therefore, the decision to purchase insurance is partly determined by the ability of the co-drivers. Shared driving experiences are likely to ensure that asymmetric information with regard to driving abilities, within the household, is minimal. The analysis was restricted to 1,462 households that garaged two vehicles, so that the insurance status of the “other” vehicles could be expressed as indicator variable equal to one if comprehensively insured and equal to zero if otherwise for a single secondary vehicle. The pair-wise correlation, with the p-value in parentheses, for the insurance status of the primary and secondary automobile was 0.338 ($<$ 0.01). Thus, the IV candidate is connected with the target variable, which economic theory suggests is likely to be endogenously determined.

The second condition, that the IV be uncorrelated with the error term, cannot be proved empirically. Nevertheless, the inability of the principal to control for the unobserved preventive effort of the agent is, theoretically, the root cause of *ex ante* moral hazard. We argue that the insurance status of the secondary vehicle has no independent effect on the probability of a RTC in the primary vehicle: comprehensive insurance purchased on the secondary vehicle has no affect on preventive effort while driving the primary vehicle. Only insurance purchased for the primary vehicle will affect the level of preventive effort invested to avoid a RTC in primary vehicle. Hence insurance status of the secondary vehicle will only be correlated with a RTC in the primary vehicle, through the insurance status of the primary vehicle. Figure 1 illustrates that the insurance status of the secondary vehicle will be correlated with a shared unobserved risk type, which intern is correlated with the insurance decision for the primary vehicle (Arrow 1985). Thus the effect of unobserved risk type, will be mediated through the endogenously determined decision to insure. Hence our IV "...is independent of all variables (including the error term) that have an influence on [a RTC] that is not mediated by [the insurance status of the primary vehicle]".

Importantly, the validity of our selected IV is not affected by either (i) who buys the insurance or (ii) who ultimately drives the primary and secondary vehicles. Firstly, the method used to purchase insurance (be that jointly or independently determined or imposed by a household dictator) will not affect the validity of the IV as long as the insurance decision embodies some private information about risk-type, held by the household. The insurance status of the primary and secondary vehicles will be correlated, thus satisfying the first condition that the IV be correlated with the endogenous variable.

Secondly, who subsequently drives each vehicle does not affect the second condition because *ex ante* the policyholder's insurance decision will reflect the driving capabilities of all drivers likely to be granted access to the vehicle. To illustrate, consider a household where the insurance decision is jointly determined, reflecting the couple's joint risk-type. Even if vehicles were randomly allocated daily, the underlying rationale for our IV still holds. Neither driver of the primary vehicle would allow the insurance status of the secondary vehicle to affect their level of preventive care, thus satisfying the second condition that the IV have no direct effect on the dependent variable. Even in households that consists of a single individual who owns and insures two vehicles, the second condition still holds: preventive effort invested while driving the primary vehicle, will not be affected by the insurance status of the secondary vehicle. Thus, our IV is uncorrelated with the error term, by assumption. As such, this IV passes both conditions that are required for the defensible application of the IV approach.

Therefore, the following bivariate probit model will be estimated with the insurance status of a secondary garaged automobile within the household instrumented for the insurance status of the primary automobile, within a sample of two-vehicle households, as follows:

$$RTC\,Car1_{it} = \alpha_0 + \alpha_1 INS\,Car1_{it} + \alpha_2 \mathbf{X}_{it} + RT^*_{it} + PE^*_{it} + \epsilon_i \tag{4}$$

$$INS\,Car1_{it} = \beta_0 + \beta_1 INS\,Car2_{it} + \beta_2 \mathbf{X}_{it} + RT^*_{it} + PE^*_{it} + \eta_i \tag{5}$$

The null hypothesis of no moral hazard is given by $H_0 : \alpha_2 = 0$. A test for residual asymmetric information is given by $H_0 : \rho = 0$. We acknowledge that, if a significant proportion of Australia's underwriters used the insurance status of other cars garaged in the household to risk-rate the primary vehicle, the $INSCar2$ should be included in the insurer's information and could not function as a valid IV. To allay this concern, we undertook a separate investigation to determine whether the five largest insurers in our dataset (jointly covering 58.7 per cent of policies in the sample) collected this variable and found they did not. Thus, we are confident that our IV is not invalidated by that argument either.

Controls and Interaction Terms

An examination of the IMRAS dataset identified three potential sources of bias. Our first concern was that recall bias might confound the results. The participants were asked if any of the following indemnifiable incidents (vehicle stolen, vehicle broken into, vehicle burnt, car-part stolen, or RTC) occurred from October 1997 to November 1999. A dichotomous variable $RTC_{1997-99}$ was constructed if the respondent indicated that an RTC had occurred. Recall bias could result if RTCs that occurred in 1997 were less frequently recalled than RTCs that occurred in 1999. However, the survey also asked (i) which incident-type occurred most recently and (ii) in which year did this incident-type occur. This information identifies four mutually exclusive RTC sub-types that we use to control for potential recall bias. The first three sub-types are comprised of those RTCs that occurred most recently in 1999 (473 RTCs), 1998 (293 RTCs), and 1997 (148 RTCs). A fourth, RTC sub-type was comprised of those residual RTCs, which were preceded by another incident-type e.g. a vehicle stolen (80

RTCs). Thus, a set of dichotomous variables identifying the four RTC sub-types were included as explanatory variables to control for possible recall bias.

A second concern was that the occurrence of other indemnifiable incidents during the two-year recall period (e.g. theft of car) might confound the analysis in unpredictable ways. For example, automobile theft may either increase the demand for insurance if the individual makes an upwards revision to his or her risk status or reduce the supply of insurance (e.g., if insurers drop claimants at the end of contract). Finkelstein and McGarry (2006) have argued that to allow for possible nonlinearities among variables, controls including interaction terms should be included. In this spirit, dichotomous variables identifying if the automobile was stolen, broken into, burnt, or car-part stolen with interaction terms identifying if an RTC occurred concurrently, were also included in our specification of the model.

Finally, our analysis compares a recursive model, which was used to analyze a complete set of households (n=4005) with the results of estimating our biprobit IV model , which was used to analyze the sub-set of two-vehicle households (n=1,462). To control for the possibility that two-vehicle households might be systematically different from other households a dichotomous variable equal to one if a two-vehicle household was included in the recursive model.

Results

We commence with a test for asymmetric information in the Australian market of automobile insurance using a bivariate probit model of the form specified by Chiappori (2000)

$$RTC_{it} = \alpha_0 + \alpha_1 \mathbf{X}_{it} + \varepsilon_{it} \tag{6}$$

$$INS_{it} = \beta_0 + \beta_1 \mathbf{X}_{it} + \nu_{it} \tag{7}$$

RTC and INS are dichotomous variables and $\mathbf{X_i}$ is a vector of variables, denoted † in Table 2, which reflect the insurer's information set. The null hypothesis of no asymmetric information is given by $H_0 : \rho = 0$. The bivariate probit model reports $\rho = 0.02$ with a p-value $= 0.74$. Thus, we cannot reject the null hypothesis. This result concurs with that of Chiappori and Salanié (2000), Dionne et al. (2013) and Finkelstein and McGarry (2006) who also report no evidence of asymmetric information with this traditional test.

Tests for Moral Hazard

Table 2, reports the results for both models which were estimated simultaneously with FIML. Our focus is the relationship between insurance and an RTC. The coefficients for the variable primary vehicle insured are 1.45 (p-value $<$ 0.01) and 0.97 (p-value $= 0.04$) in the recursive and bivariate probit models, respectively. Thus, the null hypothesis of nil *ex ante* moral hazard in the market for comprehensive vehicle insurance are rejected in both models. While a statistical correlation can never prove causation these results suggest that conditional upon the insurer's information set, the purchase of insurance is correlated with an increased probability of an RTC.

Table 2: Recursive model and bivariate probit model with one IV

	Recursive (n=4005)				Biprobit with IV (n=1462)			
	RTC 1997-99		Primary vehicle insured		RTC 1997-99		Primary vehicle insured	
Variables	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*
Primary vehicle insured	1.45	<0.01	n.a	n.a	0.97	0.04	n.a.	n.a.
Secondary vehicle insured	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.76	<0.01
RTC History								
RTC 1994-97 †	n.a.	n.a.	0.14	0.08	0.46	<0.01	-0.24	0.15
Driver Characteristics								

	Recursive (n=4005)				Biprobit with IV (n=1462)			
	RTC 1997-99		Primary vehicle insured		RTC 1997-99		Primary vehicle insured	
Variables	Coef.	p	Coef.	p	Coef.	p	Coef.	p
Aged 25 to 34 years †	-0.16	0.32	0.12	0.44	-0.17	0.64	0.77	0.02
Aged 35 to 44 years †	-0.03	0.87	0.05	0.77	-0.42	0.30	0.72	0.06
Aged 45 to 54 years †	-0.06	0.78	0.17	0.41	-0.64	0.16	0.59	0.18
Aged over 55 years †	-0.14	0.55	0.29	0.19	-0.69	0.16	0.78	0.11
Male †	0.02	0.84	-0.27	<0.01	0.00	1.00	-0.08	0.58
Metropolitan †	0.10	0.20	0.08	0.27	-0.03	0.83	0.19	0.21
Other driver < 25 years †	0.25	0.01	-0.01	0.95	0.28	0.17	0.16	0.44
Private registration †	-0.05	0.72	0.17	0.29	-0.27	0.28	-0.25	0.49
SES 2nd quartile †	-0.06	0.48	0.13	0.13	0.08	0.64	0.00	0.99
SES 3rd quartile †	-0.07	0.49	0.19	0.06	0.14	0.45	-0.17	0.36
SES 4th quartile †	-0.15	0.15	0.29	0.01	0.24	0.24	0.14	0.50
Licensed 6 to 10 years †	-0.43	0.01	0.35	0.02	-0.58	0.10	0.56	0.08
Licensed 11 to 15 years †	-0.51	0.01	0.59	<0.01	-0.63	0.11	0.35	0.33
Licensed 16 to 20 years †	-0.71	<0.01	0.75	<0.01	-0.34	0.39	0.66	0.07
Licensed 21 to 25 years †	-0.68	<0.01	0.70	<0.01	-0.27	0.52	0.56	0.16
Licensed > 25 years †	-0.71	<0.01	0.69	<0.01	-0.08	0.86	0.71	0.10
Income $20,000-$39,999	0.21	0.16	0.08	0.50	-0.02	0.94	0.03	0.91
Income $40,000-$59,999	0.11	0.49	0.23	0.08	0.08	0.82	0.30	0.30
Income $60,000-$79,999	0.20	0.23	0.16	0.26	0.08	0.81	0.18	0.56
Income $80,000-$99,999	-0.04	0.85	0.60	0.00	-0.28	0.48	0.94	0.02
Income $100,000-$149,999	0.25	0.18	0.13	0.46	0.11	0.77	0.08	0.83
Income > $150,000	0.20	0.36	0.42	0.07	-0.12	0.79	0.53	0.30
Income not divulged	0.09	0.53	0.11	0.36	0.10	0.77	0.44	0.13
Profession lower white	-0.15	0.12	-0.17	0.08	-0.33	0.05	-0.02	0.94
Profession upper blue	-0.16	0.11	-0.11	0.27	-0.24	0.22	-0.34	0.07
Profession lower blue	-0.13	0.44	-0.17	0.29	-0.56	0.15	-0.05	0.90
Profession home duties	-0.21	0.11	-0.24	0.05	-0.64	0.01	-0.19	0.40
Profession student	0.01	0.96	-0.26	0.12	-0.05	0.90	-0.30	0.39
Profession retired	-0.27	0.06	0.24	0.10	-0.49	0.08	0.16	0.58
Profession unemployed	0.02	0.93	-0.52	0.02	-0.16	0.73	-0.80	0.05
Profession not divulged	-0.11	0.71	-0.51	0.04	-5.40	1.00	-1.52	0.00
Vehicle's Characteristics								
6-cylinder vehicle †	0.03	0.74	0.05	0.61	0.06	0.74	0.26	0.14
8-cylinder vehicle †	-0.05	0.79	-0.15	0.33	0.03	0.93	-0.27	0.38
Make GM Holden †	0.02	0.83	-0.12	0.23	0.01	0.95	-0.23	0.24
Make Toyota †	0.22	0.06	0.15	0.19	0.19	0.38	0.27	0.20
Make Mitsubishi †	0.12	0.38	0.00	0.99	0.09	0.73	0.20	0.43
Make Asian †	0.21	0.06	-0.08	0.48	0.30	0.14	-0.11	0.58
Make European †	0.05	0.75	0.08	0.61	-0.25	0.42	-0.06	0.86
Body-type Commercial †	0.10	0.45	-0.18	0.14	0.04	0.87	-0.34	0.12
Body-type 4 WD †	-0.19	0.16	-0.08	0.55	-0.47	0.06	-0.24	0.38

	Recursive (n=4005)				Biprobit with IV (n=1462)			
	RTC 1997-99		Primary vehicle insured		RTC 1997-99		Primary vehicle insured	
Variables	Coef.	p	Coef.	p	Coef.	p	Coef.	p
Body-type Sports vehicle †	-0.14	0.50	0.38	0.13	0.08	0.84	0.41	0.38
Vehicle age 3 to 7 years †	0.01	0.93	-0.07	0.58	0.06	0.72	-0.43	0.10
Vehicle age 7 to 12 years †	0.25	0.02	-0.52	<0.01	0.33	0.10	-0.67	0.01
Vehicle age > 12 years †	0.29	0.07	-1.06	<0.01	0.05	0.87	-1.22	<0.01
Value $2,001-$6,000 †	-0.22	0.06	0.43	<0.01	-0.43	0.10	0.81	<0.01
Value $6,001-$10,000 †	-0.48	<0.01	0.79	<0.01	-0.38	0.18	1.19	<0.01
Value $10,001-$16,000 †	-0.52	<0.01	0.95	<0.01	-0.64	0.06	1.70	<0.01
Value $16,001-$25,000 †	-0.43	0.01	0.90	<0.01	-0.17	0.63	1.40	<0.01
Value > $25,000 †	-0.44	0.02	1.01	<0.01	-0.32	0.41	2.33	<0.01
Controls and interaction terms								
Most recent RTC 1998	7.99	1.00	-0.08	0.52	8.29	1.00	-0.08	0.72
Most recent RTC 1997	8.03	1.00	0.24	0.21	12.20	1.00	1.36	0.03
Most recent event not RTC	0.46	1.00	0.09	0.77	0.23	1.00	1.49	0.07
Vehicle stolen	-5.26	1.00	-0.07	0.77	-5.54	1.00	-0.68	0.16
Vehicle broken into	-5.62	1.00	-0.09	0.52	-5.94	1.00	-0.03	0.90
Vehicle burnt	-4.99	1.00	-1.03	0.16	-5.49	1.00	3.52	1.00
Hail damage	-5.20	1.00	-0.21	0.37	-6.31	1.00	-0.07	0.89
Other event	-6.40	1.00	5.92	1.00	-5.51	1.00	5.92	1.00
Vehicle part stolen	-6.44	1.00	6.88	1.00	colinear	colinear	colinear	colinear
Vehicle stolen * RTC	12.50	1.00	0.07	0.87	12.76	1.00	0.06	0.96
Vehicle broken into * RTC	13.02	1.00	-0.18	0.49	13.33	1.00	-0.87	0.09
Vehicle burnt * RTC	-2.36	1.00	8.15	1.00	colinear	colinear	colinear	colinear
Hail damage * RTC	12.80	1.00	0.12	0.76	13.89	1.00	0.44	0.64
Two-vehicle household	0.05	0.44	0.05	0.44	n.a.	n.a.	n.a.	n.a.
Constant	-1.35	<0.01	-0.29	0.30	-0.75	0.21	-1.06	0.10
rho(S.E) = -0.57(0.286)								

Note: The reference individual was aged 18 to 24 years, with less than 5 years licensure and lived in a postcode with the lowest SES. He/she drove a 4-cylinder, Ford sedan that was less than 3 years old and valued at less than $2000.

The statistically significant coefficients for both of the specifications reported in Table 2 are generally quite similar. In both models vehicle owners are risk-averse; vehicle value is positively correlated with comprehensive insurance decisions and vehicle age is negatively correlated with comprehensive insurance.

While driver age is not correlated with an RTC or insurance, years-licensed –a closely related construct– is strongly correlated with the dependent variables in each model. Vehicle value and vehicle age are correlated with an RTC. Some differences between the models do exist, though. For example, SES is correlated with an RTC in the recursive model but not in the bivariate probit model with one IV, possibly due to the smaller sample size for the latter specification. Note that the income and occupation-type variables that were included to proxy for policyholder information, which may be observable to the insurer but not otherwise included in our dataset, are weakly correlated with RTC and Insurance. The inclusion of controls for recall bias and concurring insurable events appears germane. In the bivariate probit model with an IV, vehicles involved in an RTC-that occurred in 1997 are more likely to have been insured and vehicles involved in an RTC concurring with theft are less likely to be insured. These variables from the survey data thus provide a thorough set of controls for risk-type and preference for insurance.

Next, we conduct a battery of tests for weak instruments. First, we report the change in the *pseudo*-R^2 between equations (8) and (9), which estimate the endogenous variable $INSCar1_{it}$, regressed on vector of exogenous variables $\mathbf{X}$, with and without the selected IV, $INSCar2_{it}$, and a likelihood ratio test on the null hypothesis of no correlation between the IV and the endogenous regressor.

$$INSCar1_{it} = \alpha_0 + \alpha_1 \mathbf{X} \tag{8}$$

$$INSCar1_{it} = \beta_0 + \beta_1 \mathbf{X} + \beta_2 INSCar2_{it} \tag{9}$$

The pseudo-R^2 increase from 0.36 to 0.41 and the LR test result ($\chi 2(1) = 45.69$), which is statistically significant at the one per cent level, suggests that the chosen IV is not weak. Secondly, following Cameron and Trivedi (2009), we first re-estimate the system of equations (4) and (5) using two-stage least squares (2SLS) with robust standard errors.[5] In our just-identified system with one endogenous regressor, we test the significance of the instrument $INS_{i,2nd\_car}$ in the first-stage regression (equation (5)). The null hypothesis that the instrument is weak is given by $H_0 : \beta_2 = 0$, and we reject the null hypothesis if the F statistic is greater than 10 (Staiger and Stock 1997). The computed an F statistic of 43.55 thus enables the null hypotheses of a weak instrument to be rejected.

We also subject our findings to two further sensitivity tests. Firstly we re-estimate both the recursive model and biprobit model with one instrumental variable with an insurer's information set that was strictly limited to those variables that Australian automobile insurers are known to collect. The coefficients for *INS* in the Recursive and Bivariate probit model with one IV were 1.58 (p-value < 0.01) and 0.88 (p-value$= 0.02$), respectively. Secondly, to ensure that our inclusion of controls and interactions terms were not confounding our results we re-estimated the biprobit model with an instrumental variable with a restricted sample of RTCs which occurred in 1999 (i.e., a maximum recall period of nine months) and no occurrence of any other insurable event (e.g. automobile stolen) which may potentially confound the results. The models were again re-estimated and the coefficients were 1.45 (p-value < 0.01) and 0.89 (p-value $=$ 0.09), respectively thus demonstrating that neither re specification materially affected our findings.

Marginal Effects

To interpret the coefficient for the endogenous binary variable, quantitatively, it is necessary to compute the estimated marginal effects. In general, the mar-

[5]Angrist (2006) has argued that 2SLS can be used to estimate binary probability models with dummy endogenous variables because linear 2SLS estimates have a robust causal interpretation that is insensitive to possible nonlinearity introduced by the dummy dependent variables.

ginal effect of any explanatory variable, which is specified in both parts of the bivariate probit model, will have both a direct and indirect effect on dependent variable (Greene 2008). The direct effect will be expressed through first equation and an indirect effect, which captures the effect on the explanatory variable that is imparted through the effect on the endogenous variable, is expressed through the second equation. When the endogenous variable is binary, simply differentiating the conditional mean function may not produce an accurate result. We therefore use a methodology recommended by Greene (1998, p. 298) to estimate the marginal effects of a binary endogenous variable. In this approach the conditional mean function is computed, first with the binary endogenous variable set to one and then set to zero and the difference taken to calculate the marginal effect:

$$E[RTC|\mathbf{X}_{\mathrm{RTC}}, \mathbf{X}_{\mathrm{INS}}, INS = 1] - E[RTC|\mathbf{X}_{\mathrm{RTC}}, \mathbf{X}_{\mathrm{INS}}, INS = 0] \tag{10}$$

$$= \Phi(\mathbf{X}_{RTC}\beta_{RTC} + \gamma) - \mathbf{\Phi}(\mathbf{X}_{RTC}\beta_{RTC}) \tag{11}$$

where Φ is the bivariate normal distribution in the cumulative distribution function, $\mathbf{X}_{RTC}$ is a vector of exogenous variables that predict *RTC,* β_{RTC} is a vector of the corresponding coefficients and γ is the estimated coefficient for *insurance.* The marginal effects of insurance were calculated by directly estimating the expected probability of a RTC, with and without insurance, using our two econometric models that controlled for endogeneity. Our results are set out in Table 3 as follows.

Three alternate specifications were considered: (a) the covariates set to their mean, (b) the covariates set to their median and (c) the covariates were set to represent a young working class male. Column 4, which is given by the difference between Columns 3 and 2, reports the marginal effect of insurance for

Table 3: Marginal effect of insurance on the probability of an RTC

Models	Probability of a RTC with insurance	Probability of a RTC without insurance	Marginal effect 1997-99	Annualized marginal effect
Covariates X set to mean				
Recursive model	0.352	0.009	0.342	0.171
Bivariate probit with IV	0.315	0.07	0.245	0.123
Covariates X set to median				
Recursive	0.106	0.0001	0.106	0.053
Bivariate probit with IV	0.016	0.001	0.014	0.007
Representative individual				
Recursive model	0.076	0.0001	0.076	0.038
Bivariate probit with IV	0.097	0.010	0.087	0.044

Note: The “representative individual” is was aged 24 to 34 years, with 6 to 10 years of licensure, male, metropolitan abode, second SES quartile, no nominated co-driver, upper blue-collar worker with an income between $40,000 and $59,999 per year. The vehicle was a privately owned, Holden, sedan, with a 4-cylinder engine, 3 to 7 years old, value of $6,000 to $10,000. They had no previous RTC in the last five years nor any other insurable event during the last two years (i.e. all interaction terms were set to zero).

the two year period 1997 to 1999 and Column 5 reports an annualized marginal effect, which is given by dividing Column 4 by 2. Table 3 reports that insurance increases the probability of a RTC from between 0.7 to 17.1 per cent, depending on the econometric model one chooses. The mid-point of these estimates is similar to the 5.9 percentage point increase in probability of an RTC reported among French drivers in an earlier working paper by Dionne et al. (2004) but less than Weisburd (2015) who attributed a 10 percentage point increase in automobile accidents to *ex ante* moral hazard.

Advantageous Selection

The tests for asymmetric information conducted by Finkelstein and McGarry (2006) included a control for advantageous selection. They argued that the variable “seat belt use” was a good proxy for risk aversion because it was correlated with risk aversion but had no direct effect on the likelihood of admission to long-term care. To capture a flavor of this analysis in our own investigation we present the following sensitivity analysis. In Australia, comprehensive insurance not only indemnifies policyholders against the cost of an RTC but also the theft of the vehicle. Australian insurers typically collect information regarding the place that the vehicle is kept overnight in order to risk-rate premia according to the probability of theft. If one accepts the argument that how the vehicle is stored does not affect the probability of an RTC *ceteris paribus,* vehicle storage can serve as a proxy for risk aversion and thus a control for advantageous selection can be included in the models. In our survey, respondents were asked how the vehicle was usually garaged. An ordinal variable vehicle garaged (=1 if parked on street, =2 if parked in carport or under apartment and =3 if garaged) was constructed. The basic premise for the construction of this variable is that increases in vehicle garaged from one through to three reflect an increase in the individual's level of risk aversion. The pair-wise correlations for *vehicle garaged* and *insured* and *vehicle garaged* and *RTC* were 0.12 (ρ<0.01) and -0.07 (ρ<0.01), respectively. These results suggest that *vehicle storage* captures some of the essence of advantageous selection, i.e. that risk-averse people who chose to store their vehicle more securely are more likely to purchase insurance but less likely to be involved in an RTC.

The Recursive and IV biprobit models were re-estimated with the additional variable *vehicle garaged* included to control for advantageous selection. The results are reported in Table 4, and illustrate two points. First, note that in the inclusion of the variable cars garaged is statistically significant in both models. In the recursive model it is positively correlated with insurance and negatively correlated with RTC while in the biprobit model with an IV it is positively correlated with insurance but not correlated with RTC. It can therefore be argued that the inclusion of this variable does capture some otherwise private information pertaining to risk aversion that could potentially confound the tests for moral hazard. Second, note that the results of the tests for moral hazard remain unchanged. Insurance remains statistically significant in both specifications and the size of the coefficients are substantively unchanged from those reported in Table 2. Thus, the tests for moral hazard described above appear reasonably robust to the possible effects of advantageous selection.

Table 4: Recursive model and bivariate probit model with one IV and a control for advantageous selection

	Recursive (n=4005)				Biprobit with IV (n=1426)			
	RTC 1997-99		Primary vehicle insured		RTC 1997-99		Primary vehicle insured	
Variables	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*
Primary vehicle insured	1.46	<0.01	n.a.	n.a.	1.03	0.02	n.a.	n.a.
Secondary vehicle insured	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	0.76	<0.01
RTC History								
RTC 1994-97 †	n.a.	n.a.	0.15	0.07	0.44	<0.01	-0.25	0.15
Advantageous Selection								
Vehicle garaged	-0.09	0.04	0.09	0.05	-0.27	0.02	0.06	0.47
Constant	-1.12	<0.01	-0.47	0.11	-0.29	0.65	-1.19	0.07

Note: Coefficients available upon request

A Falsification Test

The data at our disposal also enabled us to subject our findings to a further falsification test: in the event an RTC was reported, respondents were asked were they able to prove that someone else was at-fault. Shavell (1979) has argued that *ex ante* moral hazard occurs when insurance causes a reduction in unobserved preventive effort. We argue that while at-fault RTCs are (partially) a function of driver effort, not-at-fault RTCs occur randomly and are generally not a function of driver effort (although they may be related to exposure). It follows therefore that evidence of *ex ant*e moral hazard should be more evident in at-fault, *vis-à-vis* not-at-fault RTCs.

The full sample which was comprised of 994 RTCs consisted of 471 RTCs where the other driver was at-fault, 432 RTCs where the other driver was not-at-fault and 91 RTCs where fault was not determined. The sample was stratified and several new variables were created for our falsification test: in the first sub-sample, a binary variable was constructed to be equal to one if the participant indicated the other driver was not-at-fault and equal to zero if no RTC occurred. The not-at-fault RTCs were then removed from the sample. Similarly, in the second sub-sample, a binary variable was constructed equal to one if the participant indicated that the other driver was at-fault and zero if no RTC occurred. The at-fault RTCs were then removed from the sample. Each sub-sample was analyzed and the results reported in Table 5. While evidence of *ex ante* moral hazard persists in the sample of at-fault RTCs 1.39 (p-value < 0.01) and 1.62 (p-value < 0.01) evidence of *ex ante* moral hazard in the sample of not-at-fault RTCs is no longer present 0.85 (p-value $= 0.26$) and 0.38 (p-value $= 0.64$), thus providing strong corroboration of the evidence of *ex ante* moral hazard that was presented in Table 2.

Conclusion

In this paper, two conventional econometric strategies were used to control for endogeneity using cross-sectional survey data. The first approach used a recursive model whereby the endogenous variable insurance is estimated using the predetermined variable $RTC_{i,t-1}$ a proxy for risk-type and a vector of observable

Table 5: Recursive model and bivariate probit model with one IV and a control for advantageous selection

	Recursive model				biprobit model with IV			
	At-fault RTC (n=2396)		Not-at-fault RTC (n=2372)		At-fault RTC (n=898)		Not-at-fault RTC (n=897)	
	coef.	*p*	coef.	*p*	coef.	*p*	coef.	*p*
primary vehicle insured †	1.39	<0.01	0.85	0.26	1.62	<0.01	0.38	0.64
RTC History								
RTC 1994-97 †	n.a.	n.a.	n.a.	n.a.	0.40	0.02	0.37	0.08
Two-vehicle household	0.08	0.34	0.08	0.38	n.a.	n.a.	n.a.	n.a.
Constant	-1.64	<0.01	-1.55	0.00	-0.99	0.11	-1.33	0.14

Note: Coefficients available upon request

covariates. This approach assumes that $RTC_{i,t-1}$ is predetermined and therefore uncorrelated with the error term. However, RTC history is a variable that is widely used by insurers to risk rate their policyholders. Therefore, the extent to which the variable will be uncorrelated with the error term depends on the degree to which reported RTCs (including minor RTCs otherwise unreported to insurance firms) capture risk type. Cognizant of the possibility that our first approach may not fully account for unobserved risk-type a second model which used the insurance status of the secondary vehicle garaged at the household as an IV for the insurance status of the primary vehicle in a sub-sample of two-vehicle household was estimated. The intuitive rationale that underpins the selection of this IV was that shared household characteristics and joint vehicle operation result in a correlation between the IV and the endogenous variable but that the insurance of the secondary vehicle cannot elicit moral hazard while driving the primary car.

An advantage of using the recursive model is that the entire dataset is analyzed, however to address the inherit endogeneity within the structural model $RTC_{i,t-1}$ is assumed to be predetermined. On the other hand, the bivariate probit model that uses an IV to address the endogeneity problem restricts the analysis to a sample of two vehicle households. Results derived using this approach could be confounded if the existence of moral hazard in insured drivers living in two-vehicle household were systematically different from drivers who do not live in a two-vehicle household. The inclusion of a binary variable *two-vehicle household* in the recursive model was not, however, found to be statistically significant with respect either to an RTC or to insurance, which allays that concern.

The coefficients for insurance in the recursive model 1.45 (p-value < 0.01) and bivariate probit model 0.97 (p-value < 0.01) both imply that *ex ante* moral hazard is evident in the market for vehicle insurance. The ability to differentiate between at-fault from not-at-fault RTCs in this dataset enabled us to subject our results to further testing. While no evidence of *ex ante* moral hazard was found in the sub-sample of not-at-fault RTCs, evidence of *ex ante* moral hazard persists in the sub-sample of at-fault RTCs. These findings are consistent with our theoretical expectations for evidence of moral hazard in each sub-sample. This falsification test provides some reassurance of the veracity of the results.

Our decision to follow the methodological lead of Finkelstein and McGarry (2006) and analyze survey data, as opposed to claims data, was one important part of our approach: both of our econometric models rely on the use of data that are unobserved by insurers and hence unavailable in claims datasets. The recursive model assumes that the RTC history is a measure of the unobserved risk-type, as it includes RTCs that did not result in claims and, as such, were probably unidentified to the insurer. The bivariate probit model utilizes the insurance status of the secondary car, a variable that is not normally observable to the insurer, as an instrument. It is difficult therefore to see how either method could be applied to claims data since all pertinent and observable data would be subsumed into the vector of covariates representing the insurer's information set. We suggest that access to survey data may provide a fertile avenue to the disentanglement of adverse selection from moral hazard in other empirical studies of hidden information in insurance markets.

References

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

Abbring, Jaap H., Pierre-André Chiappori and Jean Pinquet (2003), 'Moral hazard and dynamic insurance data', *Journal of the European Economic Association* **1**(4), 767–820.

Angrist, J. (2006), 'Instrumental variables methods in experimental criminological research: What, why and how', *Journal of Experimental Criminology* **2**(1), 23–44.

Arrow, K. J. (1985), The economics of agency, *in* J. W.Pratt and R.Zeckhauser, eds, 'Principals and agents: The Structure of Business', Harvard Business School Press, Boston, chapter 2, pp. 37–51.

Australian Bureau of Statistics (2006), 'Census of population and housing: Socio-economic indexes for areas (seifa), australia - data cube only, 2001'. **URL:** *http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage /2033.0.55.0012001?OpenDocument*

Cameron, A. and P. Trivedi (2009), *Microeconometrics using Stata*, Stata Press, College Station, Texas.

Chiappori, Pierre-André (1999), Asymmetric information in automobile insurance: An overview, *in* G.Dionne and C.Laberge-Nadeau, eds, 'Automobile insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation', Kluwer Academic Publishers, Boston, chapter 1, pp. 1–11.

Chiappori, Pierre-André (2000), Econometric models of insurance under asymmetric information, *in* G.Dionne, ed., 'Handbook of Insurance', Kluwer Academic Publishers, Boston, pp. 365–393.

Chiappori, Pierre-André and Bernard Salanié (2000), 'Testing for asymmetric information in insurance markets', *The Journal of Political Economy* **108**(1), 56–78.

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

Cohen, Alma and Peter Siegelman (2010), 'Testing for adverse selection in insurance markets', *Journal of Risk and Insurance* **77**(1), 39–84.

Dionne, Georges (2013), The empirical measure of information problems with emphasis on insurance fraud and dynamic data, *in* G.Dionne, ed., 'Handbook of Insurance', 2 edn, Springer, New York, chapter 15.

Dionne, Georges, Christian Gourieroux and Charles Vanasse (2001), 'Testing for evidence of adverse selection in the automobile insurance market: A comment', *The Journal of Political Economy* **109**(2), 444–453.

Dionne, Georges, Mélissa La Haye and Anne-Sophe Bergéres (2010), *Does asymmetric information affect the premium in mergers and acquisitions? Working paper 10-03*, Canada Research Chair in Risk Management, HEC Montreal.

Dionne, Georges, Pascal St-Amour and Vencatachellum Désiré (2009), 'Asymmetric information and adverse selection in mauritian slave auctions', *The Review of Economic Studies* **76**, 1269–1295.

Dionne, Georges, Pierre-Carl Michaud and Maki Dahchour (2004), *Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal Evidence from France (Canada Research Chair in Risk Management Working Paper No. 04-05).*
URL: *http://ssrn.com/abstract=583063 or doi:10.2139/ssrn.583063 (Accessed June 2010)*

Dionne, Georges, Pierre-Carl Michaud and Maki Dahchour (2013), 'Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from france', *Journal of the European Economic Association* **11**(4), 897–917.

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

Greene, W. H. (1998), 'Gender economics courses in liberal arts colleges: Further results', *The Journal of Economic Education* **29**(4), 291–300.

Greene, W.H. (2008), *Econometric analysis*, Prentice Hall, Upper Saddle River, N.J.

Heckman, James J. and George J. Borjas (1980), 'Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence', *Economica* **47**(187), 247–283.

Hölmstrom, B. (1979), 'Moral hazard and observability', *The Bell Journal of Economics* **10**(1), 74–91.

Israel, Mark (2004), *Do we drive more safely when accidents are more expensive? Identifying Moral Hazard from Experience Rating Schemes (Mimeo)*, Kellogg School of Management, Northwestern University.

Marshall, John M. (1976), 'Moral hazard', *The American Economic Review* **66**(5), 880–890.

Mirrlees, J. A. (1999), 'The theory of moral hazard and unobservable behaviour: Part i', *The Review of Economic Studies* **66**(1), 3–21.

Pauly, Mark V. (1974), 'Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection', *The Quarterly Journal of Economics* **88**(1), 44–62.

Pearl, Judea (2009), *Causality Model, Reasoning and Inference*, Cambridge University Press, New York.

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

Rowell, D. (2011), Moral hazard: empirical evidence in the Australian market for automobile insurance, PhD thesis, School of Economics, The Univeristy of Queensland.

Shavell, Steven (1979), 'On moral hazard and insurance', *The Quarterly Journal of Economics* **93**(4), 541–562.

Spindler, Martin & Winter, Joachim & Hagmayer Steffen (2014), 'Asymmetric information in the market for automobile insurance: Evidence from germany', *The Journal of Risk and Insurance* **81**(4), 781–801.

Staiger, D. and J. Stock (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

Su, Liangjun & Spindler, Martin (2013), 'Nonparametric testing for asymmetric information', *Journal of Business & Economic Statistics* **31**(2), 208–225.

Weisburd, Sarit (2015), 'Identifying moral hazard in car insurance contracts', *The Review of Economics and Statistics* **97**(2), 301–313.

Winter, Ralph A. (1992), Moral hazard and insurance contracts, *in* G.Dionne, ed., 'Contributions to insurance economics', Kluwer Academic Publishers, Boston, chapter 3, pp. 61–96.

Wooldridge, Jeffrey M. (2000), *Introductory econometrics: A Modern Approach*, South-Western College Publishing, Mason, Ohio.

Zavadil, Tibor (2014), 'Do the better insured cause more damage? testing for asymmetric information in car insurance', *The Journal of Risk and Insurance* .