

# Asymmetric Information on Risky Behaviour: Evidence from the Automobile Insurance Market

Sara Forsstedt\*

SAMT, VTI, Box 920, Borlänge 781 29, Sweden.

This paper examines the Swedish automobile insurance market by accounting for policyholders' private information on risky behaviour in terms of major and minor traffic violations. Two approaches are used: A positive correlation test and a test where private information is used explicitly. The results show that there is a positive correlation, which is not affected when including private information in the regression, that policyholders with private information on risky behaviour are *less* likely to purchase full coverage, and that speeders follow a varying pattern. The conclusion is that it is favourable to use private information explicitly when asymmetric information is considered, rather than base the conclusion solely on the correlation test.

*The Geneva Papers* (2014) 39, 104–145. doi:10.1057/gpp.2013.19

**Keywords:** Insurance data; asymmetric information; adverse selection; moral hazard

Article submitted 28 March 2011; accepted 26 April 2013; published online 31 July 2013

---

## Introduction

Asymmetric information has long been alleged to cause inefficiencies in insurance markets. However, research has shown that the empirical findings regarding automobile insurance markets are ambiguous regarding the core prediction that individuals with extensive coverage are more likely to be high risks for the insurer.<sup>1</sup> Most previous papers have interpreted the absence of a significant coverage–risk correlation to mean that the contract-relevant information asymmetry is successfully handled by the insurer. Other explanations such as the absence of useful private information and policyholders' inability to act on private information have also been suggested. In addition, previous research has found that both higher and lower risks demand insurance, which in aggregate suggests that those with more insurance are not

---

\* Formerly Sara Arvidsson

<sup>1</sup> Note that the measurement of risk is *the expected reimbursement*, that is, the risk associated with the insurance company. The distinction is rather interesting, especially in Sweden, where the premiums are generally low for private insurance (the category into which automobile insurance falls), and much of the cost associated with accidents is borne by the social insurances (which are financed via taxes). From a social perspective, it is important to reduce the accident risk. A change in the deductible may only reduce the incentive to report the accident to the insurer, but the accident probability may be unaffected. That is, a potential caveat is that the risk to the insurer is reduced (expected reimbursement), but not to society (the expected cost of an accident). See for instance Koufopoulos (2007) for details on relevant measure of risk.

higher risks, and that the market can suffer from inefficiencies even in the absence of a positive correlation between insurance coverage and risk occurrence.<sup>2</sup> Cohen and Siegelman<sup>3</sup> argue that rather than trying to resolve the question of the existence of information asymmetries once and for all, future work should try to identify circumstances under which one may expect to find evidence of relevant information asymmetry. Since market heterogeneity may play an important role, it is difficult to generalise across insurance markets and between countries. It is furthermore reasonable that the correlation structure differs across subsets of policyholders. One reason is that the information asymmetry between the insurer and the policyholder is not constant and may therefore differ across groups, for example, between new and long-term policyholders.

This paper contributes to the empirical coverage–risk literature by testing for information asymmetries with the explicit use of policyholders’ private information on risky behaviour (traffic violations). The analysis is based on a rich data set of automobile insurance policies, provided by one of Sweden’s largest insurance companies. Private information is represented by observed traffic safety violations in terms of on-the-spot fines and convictions for traffic offences. This information is unobserved and unattainable by law for Swedish insurance companies and therefore is not used in setting premiums. Additionally, because Swedish insurers are not allowed to share claim history (and other pricing characteristics), it is not possible for them to observe risky behaviour among new customers via previous claims other than by the policyholders’ self-reported claim history. This implies that high-risk drivers have incentives to switch to a new insurer and under-report claim history after a claim is made with the current insurer.<sup>4</sup> Hence, the information asymmetry is likely to be larger for new policyholders and decrease over time as the insurer receives more observations on the policyholder.

This study differs from previous studies in four ways: First, we include policyholders’ private information on risky behaviour (traffic violations), using a sample of new policyholders for whom the information asymmetry is likely to be largest. The advantage is that we are able to directly observe the effect of private information on risky behaviour, which implies that our conclusions are not all dependent on the existence of a coverage–risk correlation. Second, we use several subgroups of new policyholders that correspond to the insurer’s group classification by age and gender, which provide more homogeneous subgroups compared to the previous literature. Third, we put a restriction on vehicle age, since it may be an important determinant of choice of coverage and how the vehicle is used. Fourth, with access to all information that the insurance company uses in setting prices, we test whether the existence of private information confirms the positive correlation between risk and coverage predicted by theory.<sup>5</sup> The risk–coverage correlation calls for a remark: A positive and significant correlation is a central prediction of both adverse selection and moral

---

<sup>2</sup> Finkelstein and McGarry (2006).

<sup>3</sup> Cohen and Siegelman (2010).

<sup>4</sup> Cohen (2005) and Arvidsson (2010).

<sup>5</sup> We gained access to all the pricing variables via one of the actuaries at the insurer Länsförsäkringar AB. Since some of them are viewed as confidential by the company, we only report the variables of interest to this study.

hazard and only suggests that the presence of adverse selection or moral hazard cannot be rejected. However, disentangling adverse selection and moral hazard as well as propitious selection and preventive actions from each other is beyond the scope of the present paper.

Two approaches are used. The first is the widely used correlation test suggested by Chiappori and Salanié.<sup>6</sup> If there exists a significant correlation between risk and coverage, the null hypothesis of no residual asymmetric information is rejected. Second, we use an approach similar to that suggested by Finkelstein and McGarry,<sup>2</sup> where the effect of private information on traffic violations (risky behaviour) is directly observed.

The results suggest that there is residual private information that is positive and statistically significant for three groups, which is unaffected when including private information in risky behaviour; hence the correlation test does not seem to capture private information on risky behaviour.

According to theory, adverse selection models predict that high risk types will purchase higher insurance coverage when policyholders have private information about their risk type. In contrast to the theoretical prediction, we find that private information that the insured has higher risk actually *decreases* the probability of full insurance coverage. One exception is risky behaviour in terms of speeding, since the speeding coefficient varies for different groups of policyholders: Younger age groups are more likely to have full insurance while older age groups are less likely to have full insurance. The varying results of speeding may be explained by “common behaviour”; that is, speeding is generally viewed as an accepted violation and may therefore not be thought of as a risky behaviour by the policyholder. Hence, the policyholder may not account for, or use, this information when purchasing insurance, which results in a varying pattern by chance. Major infractions and traffic offences other than speeding tend to be unacceptable, and, according to psychological research, individuals with these infractions perceive themselves as less risky.<sup>7</sup> In that sense it is rational to purchase less insurance, which can explain why policyholders with higher risk (other than speeding) have less than full insurance. Without information on the policyholders’ perceived risk, it is difficult to fully understand why speeding tends to reflect a varying pattern and why higher risks are essentially less likely to be fully insured. Further research could benefit from survey data on infractions, perceived risk and insurance. Moreover, previous research has established that traffic violations have a significant effect on crash rates.<sup>8</sup> Our results weakly support this since few of the coefficients of risky behaviour are significant. However, if high-risk individuals tend to have less coverage, they will have even less incentive to report a claim to their insurer.

All in all, our results may explain why there previously has been ambiguity as to whether or not empirical findings support the presence of adverse selection and/or moral hazard in the automobile insurance market. If high-risk drivers essentially are less prone to have extensive insurance, we cannot expect to find a positive correlation predicted by theory. We conclude that it is preferable to study private information

---

<sup>6</sup> Chiappori and Salanié (2000).

<sup>7</sup> Forward *et al.* (2000) and Forward (2006, 2008).

<sup>8</sup> Åberg (1998).

explicitly, since it is possible to directly observe how the market is affected by asymmetric information, rather than trying to interpret (the lack of) a significant correlation.

The rest of the paper is organised as follows. The next section provides a summary of prior theoretical and empirical research with a focus on insurance markets. The section also contains information on insurance coverage and risk classification in the Swedish automobile insurance market. The subsequent section describes the empirical approach in terms of data and econometrics in more detail. The penultimate section presents the results and the last section concludes the paper.

## Background

### *Previous work*

Ever since the 1970s, the theoretical research on asymmetric information has developed at a quick pace. The prediction is that asymmetric information is a fundamental problem in most insurance markets: Policyholders are heterogeneous in risk and this risk level is private (hidden) information that is important for the contract, but unobservable to the insurer. According to the standard interpretation, the asymmetry results in a situation where high-risk individuals are associated with extensive insurance coverage, which predicts a positive correlation between *ex post* risk and extensive coverage.<sup>9</sup> Several studies, both theoretical and empirical, have also suggested the possibility of propitious (favourable) selection. These individuals have a high demand for insurance and are good risks *ex post*, and this selection predicts a negative correlation between insurance coverage and *ex post* risk occurrence.<sup>10,11</sup>

Empirical research on asymmetric information lagged behind and did not significantly evolve until the 1990s. As discussed by Chiappori and Salanié,<sup>12</sup> data from insurers are well suited for studies of asymmetric information, because they record choice of coverage and outcome (claim or not), as well as many characteristics of policyholders. Empirical studies have used data from different insurance markets and found evidence of a coverage–risk correlation.<sup>13</sup> Yet, empirical tests on property/liability insurance, where automobile insurance data have been used, do not provide any strong evidence of information asymmetries that affect the level of risk in the contract.<sup>14</sup> Three early studies suggested the presence of a positive correlation, but these were later criticised as unreliable.<sup>15</sup> Dionne *et al.*<sup>16</sup> suggest that the insurers'

---

<sup>9</sup> Rothschild and Stiglitz (1976), Bolton and Dewatripont (2005) and Salanié (2005).

<sup>10</sup> Hemenway (1990), de Meza and Webb (2001), Finkelstein and McGarry (2006) and Fang *et al.* (2008).

<sup>11</sup> De Donder and Hindriks (2009), however, show that, under some mild regularity assumptions, this prediction still does not imply a negative correlation between risk and insurance coverage in equilibrium. The reason is that there is a moral hazard effect; after obtaining insurance, the policyholder becomes less risk averse since most of the economic risk is transferred to the insurer.

<sup>12</sup> Chiappori and Salanié (1997).

<sup>13</sup> See for example Cutler (2000) and Finkelstein and Poterba (2004).

<sup>14</sup> See Chiappori and Salanié (2003) for a review.

<sup>15</sup> Dahlby (1983, 1992) and Puelz and Snow (1994).

<sup>16</sup> Dionne *et al.* (2001).

information set is sufficient if non-linear effects, not considered by Puelz and Snow, are taken into account. A sufficient risk classification implies that there is no residual adverse selection in each risk class, since groups are homogeneous in risk. Neither do they find evidence of information asymmetries using French automobile insurance data. To overcome previous difficulties with estimation, Chiappori and Salanié<sup>6</sup> (hereafter C&S) introduced a simple and general test of the presence of asymmetric information. When this test was applied to a homogeneous sample of inexperienced drivers in the French automobile insurance market, no significant correlation was found.

Cohen<sup>17</sup> argues that young drivers may not have private information since they have not learned their own risk type; that is, when policyholders learn their risk type, they develop private information. The study takes several implications of the previous critique into account, and uses a rich data set of the first five years of one start-up insurer in Israel. When applying the C&S correlation test to policyholders with less than three years of driving experience, the results are confirmed, since no significant correlation is found. However, for a group with more than three years of driving experience, Cohen finds a significant correlation between risk and coverage. The main conclusion, as drawn from results indicating that low deductible contracts are associated with more claims, is that the market is characterised by asymmetric information.

Finkelstein and McGarry<sup>2</sup> further consider the policyholder's private information on risk in the long-term medical care insurance market. Their findings indicate that two types of individuals buy insurance: Those with private beliefs that they are high risks and those with a strong taste for insurance. *Ex post*, the former are a higher risk and the latter a lower risk to the insurer. They conclude that, in aggregate, individuals with more insurance are not higher risks, and that an equilibrium with multiple forms of private information is unlikely to be efficient relative to the first best. One reason is that premiums may not be actuarially fair.

This paper differs from the previous literature mainly because we include policyholders' private information on risky behaviour (traffic violations) in the analysis. Since pricing characteristics and previous claims are not shared between insurers, we expect that the market will suffer from this asymmetric information if policyholders use their private information in their insurance decision. Another advantage is that we, via the insurance company, received access to all information that the insurance company uses in setting prices. We also put a restriction on vehicle age since the value likely affects the insurance decision. New vehicles generally do not need full coverage since they are often covered by a warranty, and old vehicles may have too low a value. Hence vehicle age will determine how much insurance to purchase. In line with previous research, we divide policyholders into homogeneous groups, but in contrast we perform the analysis on smaller subgroups with respect to age and gender (as used by the insurer).

---

<sup>17</sup> Cohen (2005).

**Table 1** Swedish automobile insurance

<i>Traffic insurance (TI)</i>	<i>Mandatory</i>	<i>Minimum coverage</i>
Limited damage insurance (LDI)	Voluntary	TI + covers glass, theft, etc.
All-risk insurance (ARI)	Voluntary	TI + LDI + covers motor vehicle damage and vehicle damage to the policyholder's vehicle when at fault, etc. <ul style="list-style-type: none"> <li>• High deductible 5,000 SEK<sup>a</sup></li> <li>• Low deductible 3,000 SEK</li> </ul>
Additional insurance	Voluntary	Assistance if the vehicle breaks down, etc.

<sup>a</sup>1 USD = 6.7 SEK.*Automobile insurance and premium pricing in Sweden*

Swedish law requires all vehicle owners to purchase traffic insurance, which is a liability insurance that covers accident damage to other drivers and their cars. Hence, the vehicle owner is equivalent to the policyholder. Note that other drivers do not have to be included in the insurance contract in order to drive the vehicle, which implies that drivers other than the policyholders are not considered when setting prices. The main reason is that Swedish automobile insurance is a property insurance, and driver and passenger costs, such as hospital care, in general, are covered by social insurance, which in turn is financed via tax.

Table 1 provides a summary of the Swedish automobile insurance policies. All-risk insurance (ARI) is the most extensive coverage on offer since it also indemnifies damages to the policyholder's own car when s/he is at fault in the claim. ARI is typically differentiated by the value of the deductible where the lower deductible provides the most extensive coverage. Additional insurance provides extra service such as a replacement car while the insured car is being repaired. The most typical comprehensive coverage in Sweden is ARI, which we focus on in this paper.

Previous studies have pointed out the importance of careful conditioning on the information set available to the insurance company. The information set is equivalent to all information that is observable and used in premium pricing by the insurance company. However, an important distinction must be made between the information set available to the insurer and the actual risk classification used in premium pricing. The information set is the basis for the actuarial prediction that results in a risk classification, and the preferred approach is therefore to condition on the companies' actuarial risk classification. The main reason is that groups of individuals with similar risk classification are considered as homogeneous by the insurer. A proper implementation of the positive correlation test therefore requires that insurance demand is analysed across homogeneous groups of individuals who are likely to face the same set of possible insurance contracts. A misspecification may result in a spurious correlation; accuracy is therefore crucial.

All information on the insurance company and how it sets prices has been obtained by interviewing one of the actuaries at the company. According to the insurer, all Swedish automobile insurance companies base their risk classification on three main categories: Risk characteristics related to the driver, the vehicle and the residential area

**Table 2** The pricing variables

<i>Vehicle</i>	<i>Description</i>	<i>Observable by the insurer?</i>
<ul style="list-style-type: none"> <li>● Vehicle year model</li> <li>● Vehicle risk classification</li> </ul>	Age of vehicle The insurer's predicted risk based on brand, motor capacity, etc. Classified information	Yes, via the licence plate Yes, constructed by the insurer
<ul style="list-style-type: none"> <li>● Kilometre class (= Annual mileage)</li> </ul>	Risk classification 1–5, based on distance driven. Most policyholders claim to drive 10,000–15,000 km/year, which correspond to risk class 2	No, but possible to verify if an accident occurs. This is rarely done
<i>Policyholder = vehicle owner</i>	<ul style="list-style-type: none"> <li>● The owner of a vehicle is obliged to purchase at least minimum coverage</li> <li>● The main user of the vehicle may not be the vehicle owner</li> </ul>	Partially observable. The insurer may receive information if another driver was involved in the accident
<ul style="list-style-type: none"> <li>● Year of birth</li> </ul>	Age of policyholder	Yes, via the personal identity number
<ul style="list-style-type: none"> <li>● Gender</li> </ul>	Female = 1, Male = 0	Yes
<i>Residential area</i> (identified via the national registration address)	The insurer's predicted risk based on number of days with icy roads, if the policyholder lives in a city or small town, etc. Classified information	Yes, constructed by the insurer from the national registration address. BUT possible for the policyholder to register at another address to lower the premium

(see Table 2). There are several variables in most of these categories, which imply that the data are fairly rich. Information that statistically affects the expected cost of offering insurance is used to establish pricing. In this way, insurers develop a risk classification that is associated with observable characteristics. The insurance contracts are thereafter divided into homogeneous groups of risk according to observable characteristics, and individuals in the same group are charged the same insurance premium. Since the 1990s, each Swedish insurer has used its own formula for determining insurance premiums. The company studied in this paper used to have a bonus-malus system, but this was gradually phased out. However, the policyholder does receive a discount for every year s/he does not make a claim.

The insurers are not allowed to share information about previous claims, so the market structure is similar in that respect to the Israeli market studied by Cohen.<sup>17</sup> The implication of not sharing claims is that policyholders may under-report their claim history when joining a new insurer in order to obtain a lower premium. This further implies that high-risk drivers have an incentive to switch insurer.

In addition, several pricing variables are based on policyholders' self-reports, for example:

- (i) Annual mileage, which consists of 1–5 risk classes. Most policyholders claim risk class 2 (10,000–15,000 km/year).
- (ii) Vehicle owner (=policyholder) vs chief user of the vehicle. A common problem in Sweden is that a parent, who generally has favourable premium ratings due to seniority, is the signed owner, while the vehicle is used by a son or daughter, who

generally has unfavourable premium ratings due to inexperience. Besides, it is often an advantage to let the woman in the household own and insure the vehicle.<sup>18</sup> This gender difference in premium rating evens out with age. The insurer cannot observe this chief user problem except when an accident, where a driver other than the owner is involved, is reported.

- (iii) Residential area, which is the national registration address. It is often more expensive to insure the vehicle in a large city; hence, registering in a smaller town means a lower premium.

It is clear that policyholders have incentives to report untruthfully in order to receive a lower premium, and that the Swedish automobile insurance market may suffer from the above information asymmetry, since it obstructs the construction of homogeneous groups and premium pricing.

## The empirical framework

### *Data*

To investigate the nature of private information, we use a rich data set that includes all the information the insurer has about its policyholders. We add data on the policyholder's risky behaviour (traffic violations), which represent the policyholder's private information. It is possible for researchers to gain access to data on traffic violations after various applications, but otherwise this information is not available, attainable or possible to observe for Swedish insurers.<sup>19</sup>

The insurer makes three main assumptions regarding the contracts. First, there is independence between contracts, the outcome for different insurance policies being independent. This implies that each contract is treated separately; an individual who owns and insures two vehicles is considered to own two contracts, one independent of the other. If the policyholder causes an accident with one of the vehicles, only the insurance contract associated with that vehicle is affected. Second, the company assumes that the cost of a claim in this period does not affect the cost in the next period (time independence). The argument is that an individual involved in an accident will drive more carefully and may be (i) less likely to have an accident in the future or (ii) a reckless driver who is always more prone to cause accidents. However, the insurer reduces the probability of claims by including a discount for each claim-free year with the company.<sup>20</sup> Third, homogeneity is assumed; an outcome with the same exposure has the same distribution within a risk group.<sup>21</sup> We therefore regard a repeated

---

<sup>18</sup> Since 21 December 2012 it is no longer possible to set prices by gender due to a change in the law.

<sup>19</sup> In order to obtain these data, the project was reviewed by the Ethical Vetting Board in Uppsala ([www.EPN.se](http://www.EPN.se)), and after approval, each authority (BRÅ and the Swedish Police) performed a secrecy review in order to decide whether or not to provide data.

<sup>20</sup> Hence, there is no strict time independence.

<sup>21</sup> There are several examples of when these conditions are violated. One example, already discussed, is untruthful reports by the policyholders, which violates homogeneity. Furthermore, if two vehicles insured by the same insurer are involved in a collision with each other, the independence between contracts could be violated.



contract as a new observation and do not consider dependency between contracts owned by the same individual.<sup>22</sup> Accordingly, since the company lacks this information about new policyholders, our subsample consists of new policyholders for whom we do not take into account the number of years without claims.

The automobile insurance data used in this study come from an automobile insurance provider with 24 regional subsidiaries located in all the counties in Sweden; its market share is approximately 32 per cent of the property insurance market. All in all, the data set contains information on 2,424,525 policy-ids and 584,425 claims, and covers three years (2006–2008). Most of the contracts are repeated and the number of observations when including those is 9,342,749.<sup>23</sup> Each observation includes all the information that the insurer has about the policyholder, vehicle and contract characteristics.

Data on the number of convictions for traffic safety violations are registered by the Swedish National Council for Crime Prevention (BRÅ). These represent major infractions such as driving while intoxicated and driving very carelessly. Data on on-the-spot fines come from the central fines register of the Swedish National Police Board (RIOB) and represents minor infractions such as speeding, running red lights, overtaking at crossings, and other offences due to risky behaviour or vehicle flaws. Since RIOB is cleared periodically, it is possible to obtain at most five years from the current year.

Fines for speeding are separated out from traffic offences, since speeding is generally viewed as a socially acceptable violation, while convictions and other traffic offences are not.<sup>24</sup> Social acceptance may affect how policyholders perceive and account for their risk, which in turn affects how much insurance they purchase.<sup>25</sup> We further separate major infractions into one conviction (=1) and more than one conviction

---

<sup>22</sup> An individual (or contract) may appear as several observations if he or she owns several cars, makes more than one claim or if any changes are made in the contract. About 25 per cent of our sample of new policyholders used in the analysis appear as two or more observations, and 11 per cent have three observations or more. We have performed a sensitivity analysis of dependency between individuals in unreported regressions. First, including only the first observation (first contract) of the individual and, second, we have cluster adjusted the standard errors with respect to the policyholder-id, and the results seem robust. Nonetheless, in this paper we consider the insurance contracts rather than the policyholders, and 22 per cent of the contracts appear as two or more observations, and 5 per cent as two observations or more. Note that, if a change is made, the contract will have a new duration and is thus a repeated contract. This implies that the only time a contract with the same date will appear as more than one observation is when more than one accident occurs (approximately 0.7 per cent). We therefore do not consider dependency between time periods, contracts and individuals.

<sup>23</sup> A note on the number of observations: Data are not truncated, which implies that we can observe policyholders back in time. Besides, there is not always only one observation per contract and year, since it is possible to change the contract during the period. If the policyholder moves or deregisters the vehicle, the risk changes and there will be a renewed, or repeated, contract and hence another observation. This implies that the total number of observations will not correspond to the number of policy-id multiplied by the number of years (2006–2008).

<sup>24</sup> According to the Swedish Transport Administration, about 55 per cent of Swedish drivers exceeded the legal speed limit speed by 5–10 per cent. See Åberg (1993) and Åberg and Rimmö (1998) and Forward (2006, 2008) and Forward *et al.* (2000) for traffic violations and acceptance and correlation with accidents.

<sup>25</sup> See Guppy (1993) for perceived risk and traffic violations.

(> 1). As we believe that relapsed criminals are higher risks, one conviction may be random, but not several. Fines for speeding will be referred to as “Speeding” and all other traffic offences as “Other traffic offences”. One conviction will be referred to as “One conviction” (=1) and more than one conviction (> 1) as “Several convictions”.

The probability of detection when committing a traffic violation is unknown, but it is reasonable to assume that a repeat offender is caught at some time. All data are matched to personal identity numbers. Data in respect of fines and convictions have been merged with the insurance and claim files by BRÅ for our project. We have also merged the insurance and claim files and cleaned the data. Appendix D provides a list of all the information that is included in each observation.

### *The subsample used and descriptive statistics*

We consider new policyholders, on whom the insurer has no previous observations, during their first year with the insurer, which provides us with a smaller subsample of 295,846 observations. Since automobile insurance is property/liability insurance, the contracts rather than the policyholders are considered.<sup>26</sup> More specifically, we sort data for policyholders who joined the insurer in 2007 and 2008, and include all contracts signed by new policyholders in 2007 and observe these contracts until they expire. For new policyholders in 2008, we observe all contracts signed in 2008 until they expire or until the end of 2008 when the data were collected. This implies that data are censored for contracts that began in 2008 and ended in 2009.<sup>27</sup> This is a general problem with insurance data, since it is possible to sign up for a one-year contract at any time in most countries.<sup>28</sup>

We further divide the policyholders into homogeneous age and gender groups that correspond to the actuarial model used during 2006–2008. This gives us 10 groups on which we perform the analysis.

We restrict our analysis to vehicles of age 3–20 years. The restriction on vehicle age is due to new vehicles generally having a motor vehicle damage warranty that corresponds to ARI. This affects the choice of purchasing more extensive coverage.<sup>29</sup> We also expect that ARI is less likely for older vehicles due to their lower value. As can be seen in Figure 1, the data confirm that the number of vehicles with ARI increases when the vehicle is three years old and decreases as the vehicle gets older. We also perform a sensitivity analysis of this restriction (Tables A1–A3).

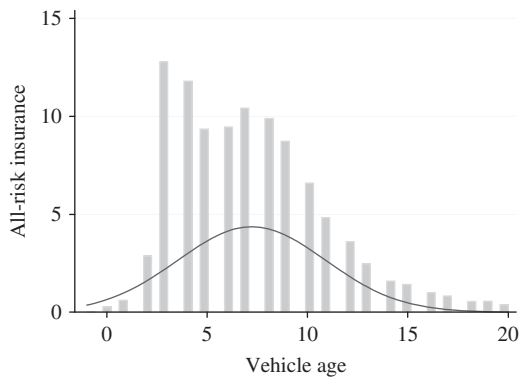
---

<sup>26</sup> Note that a policyholder may have several contracts that are viewed as different risks by the insurer. An example is a policyholder who insures vehicles of different makes.

<sup>27</sup> We performed a sensitivity analysis of the correlation test by using only new policyholders for 2007, for whom we observed the whole lifespan of the contracts; the results can be found in Table B1 in Appendix B.

<sup>28</sup> Hence, the contracts do not start in January and end in December, but can go from June to June, September to September, etc.

<sup>29</sup> Approximately, 15 per cent of policyholders tend to have ARI on vehicles below three years of age. One reason is that the deductible for the warranty is very high for some vehicle makes while some do not come with a warranty.



**Figure 1.** All-risk insurance and vehicle age.

*Note:* Vehicle age is  $-1$  to 20 years, a negative age is possible in cases where the policyholder owns a vehicle of the latest vehicle year model.

Table 3 provides descriptive statistics of the private information variables for those choosing full coverage and those that choose less than full coverage for both the whole sample and the subset of new policyholders. The table shows that individuals with less than full coverage tend to have more convictions and traffic offences. Until recently, age was not a restriction when owning and insuring a vehicle, implying that in our data very old individuals and small children are included as chief users. Individuals under 18 years of age are sorted out from the analysis, since they are too young to have a driving licence. Some individuals appear to be too old to drive; there are 168 observations of individuals aged 90 years and over in the sample of new policyholders. We do not exclude them from the analysis, since there is no upper restriction on driver age in Sweden.

Further, the maximum number of convictions is very high in all groups. Less than 1 per cent of those with convictions have more than 10 convictions, and the mean of convictions is 1.6. Those with inordinately many convictions are likely to be individuals known, and often checked, by the police. These individuals are probably not allowed to drive, but this is not a restriction on insuring a vehicle. All in all, a higher share of those with less than full insurance coverage have fines or convictions compared with those with full insurance.

### *Econometric approach*

First, we use the standard positive correlation test by C&S to examine the relationship between insurance coverage and *ex post* risk occurrence where the policyholder is held fully or partially responsible in the reported claim. That is, the policyholder fully, or partially, caused the accident in the claim. The insurer uses several degrees of causation such as fully, partial or slight. The main reason for using at-fault claims is that the insurer does not include information on other drivers than the vehicle owner (=policyholder) in the contract. Taking into account all claims, including claims involving drivers other than the policyholder, may generate a positive correlation. The

**Table 3** Policyholders with full coverage, less than full coverage and new policyholders with full coverage and less than full coverage

	<i>All policyholders with full coverage</i>				<i>All policyholders without full coverage</i>				<i>New policyholders with full coverage</i>				<i>New policyholders without full coverage</i>			
	<i>Per cent</i>	<i>Min</i>	<i>Max</i>	<i>Std</i>	<i>Per cent</i>	<i>Min</i>	<i>Max</i>	<i>Std</i>	<i>Per cent</i>	<i>Min</i>	<i>Max</i>	<i>Std</i>	<i>Per cent</i>	<i>Min</i>	<i>Max</i>	<i>Std</i>
Average year of birth		1904	2004	14.84		1898	2004	15.39		1909	1990	15.79		1908	1990	15.56
One conviction	7.1				8.4				6.1				6.8			
Several convictions	2.2				4.1				2.3				5.2			
Total number of convictions		0	82	0.56		0	136	1.03		0	61	0.75		0	117	1.92
Traffic offences	0.3	0	19	0.72	8.7	0	38	0.92	6.8	0	18	0.86	11.4	0	30	1.14
Speeding	0.5	0	8	0.67	12.8	0	8	0.70	11.9	0	7	0.69	13.0	0	7	0.72
Observations		4,144,217				5,198,532				112,531				183,315		

*Note:* The variable one conviction takes the value 1 if the policyholder (vehicle owner) has precisely one conviction (= 1), the variable “Several convictions” takes the value 1 if the policyholder has more than one conviction (> 1). Convictions represent major infractions, while fines represent minor infractions. The variable “Traffic offences” takes the value 1 if the policyholder has at least one fine for a traffic offence such as running red lights, etc. Finally, the variable “Speeding” takes the value 1 if the policyholder has at least one fine for speeding.

reason is that the variable *additional drivers* is not included among the control variables.<sup>30</sup>

We apply the bivariate probit model suggested by C&S to test for residual asymmetric information:

$$c_i = 1(\mathbf{X}\beta_1 + \varepsilon_i > 0), \quad (1)$$

$$y_i = 1(\mathbf{X}\delta_1 + \eta_i > 0), \quad (2)$$

$$(i = \text{contract}).$$

The dependent variable of Eq. (1) represents the choice of a particular contract;  $c_i=1$  if the policyholder has the highest possible coverage, that is, ARI with low deductible (3,000 SEK) and  $c_i=0$  if less coverage is bought (ARI with high deductible (5,000 SEK), limited damage insurance or traffic insurance).<sup>31</sup> The dependent variable of Eq. (2) represents an at-fault claim;  $y_i=1$  if the policyholder has a claim where s/he is fully, partially or slightly at fault,  $y_i=0$  if the policyholder is not at fault or if no claim is made.  $\mathbf{X}$  is a vector of covariates that is included to control for the risk classification used by the insurer in 2006–2008.

C&S argue that the policyholder's probability of owning a certain contract depends on the risk classification  $\mathbf{X}$  and some random shock  $\varepsilon_i$ . In a similar manner, for any  $\mathbf{X}$ , the occurrence of an accident at-fault also depends on some random shock  $\eta_i$ . The error terms are aimed at capturing any residual heterogeneity across agents when the risk classification has been taken into account. The variable of interest is the correlation between the error terms ( $\rho$ ). If  $\rho > 0$ , there is an indication of adverse selection and/or moral hazard since, conditional on risk classification, the choice of a contract and the occurrence of an accident are not independent: Contracts with more complete coverage predict a higher probability of an *ex post* risk.

In the next step of our analysis we study the effect of private information head on, by using an approach introduced by Finkelstein and McGarry.<sup>2</sup> This approach suggests that the null hypothesis of symmetric information can be rejected if, conditional on the information used by the insurer in setting prices, the econometrician observes some other characteristics of the individual that are correlated with both insurance coverage and *ex post* risk occurrence. This characteristic must be unknown, or unused, by the insurer. Finkelstein and McGarry argue that this approach provides a more robust test for asymmetric information compared to the correlation test. The reason is that it includes variables that represent the policyholder's private information, which opens up the possibility of directly observing the effect of private information. In our approach, we include the policyholders' private information about risky behaviour, which makes it possible to study the effect of private information on demand for insurance and outcome (at-fault claim or not). The null hypothesis of no

<sup>30</sup> We perform a sensitivity analysis of this restriction in the Appendices including all at-fault claims and all claims, respectively.

<sup>31</sup> 3,000 SEK correspond to approximately US\$429 and 5,000 SEK to US\$715 with an exchange rate of 6.99 SEK/USD.

residual asymmetric information is rejected if, conditional on  $\mathbf{X}$ , private information about traffic behaviour is correlated with both insurance coverage and *ex post* risk occurrence. We test the effect of private information by estimating the following probit models:

$$c_i = 1(\mathbf{X}\beta_1 + D\beta_2 + \varepsilon_i > 0), \quad (3)$$

$$y_i = 1(\mathbf{X}\delta_1 + D\delta_2 + \eta_i > 0), \quad (4)$$

( $i = \text{contract}$ ).

The added information compared to Eqs. (1) and (2) is four indicator variables that take the value 1 if the policyholder has at least one fine for speeding, at least one fine for other traffic offences, one conviction for traffic safety violation (=1), and more than one conviction (> 1) for traffic violation, respectively.

The coefficients of interest in Eqs. (3) and (4) are  $\beta_2$  and  $\delta_2$ . From them we can conclude whether the policyholder's private information on risky traffic behaviour has any effect on choosing extensive coverage, and/or the probability of being at fault in a claim. A positive correlation prediction is that  $\beta_2 > 0$  and  $\delta_2 > 0$ , which imply that violations of traffic law regulations are associated with more coverage and culpa in claims.

The variables in  $\mathbf{X}$  in all regressions are age and gender of policyholder, vehicle age, kilometre class, vehicle risk classification and residential area risk classification. We also apply the analysis to the age and gender groups used by the insurer in the actuarial model for 2007 and 2008.<sup>32</sup> Note that all coefficients are not reported here since the risk classification variables and summary statistics are the insurance company's classified information.

## Results

### *Replication of previous studies*

As discussed earlier, Cohen<sup>17</sup> found a statistically significant correlation for the more-experienced driver group. We replicate these findings by dividing the policyholders into similar groups; the results are reported in Table 4. Our approach differs in that we focus on at-fault claims and more extensive coverage compared to Cohen, who is concerned with whether low-deductible policyholders are associated with more claims. Our data do not allow such an approach, since we do not have information about indemnities; thus, we are not able to drop claims that are lower than the highest deductible when comparing the number of claims. Furthermore, since driving experience is not used in the risk classification, the data do not contain information about it. We therefore use a proxy for driving experience by considering the age group

---

<sup>32</sup> Note that the groups are differentiated according to age and gender, except for two groups that are only differentiated by age.

**Table 4** Correlation test for policyholders with different driving experience

	<i>Driving experience &lt; 3 years</i>	<i>Driving experience &gt; 3 years</i>
Correlation coefficient ( $\rho$ )	0.032 (0.042)	0.037*** (0.009)
Log-likelihood	-5,700.0135	-222,277.62
Prob > $\chi^2$	0.0000	0.0000
<i>N</i>	15,882	340,501

*Notes:* Dependent variables are at-fault claims and all-risk insurance with the low deductible (the highest possible insurance coverage) on insurers' risk classification. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

18–20 years to have less than three years of driving experience and older drivers to have more than three years of driving experience. Group one has no statistically significant correlation between risk and coverage, which confirms the results of both Cohen<sup>17</sup> and C&S. The second group, which corresponds to drivers with more than three years of driving experience, has a statistically significant correlation. Our results confirm the findings of Cohen in that we find a statistically significant correlation between risk and coverage. The correlation coefficient is low for both groups, and the reason why the correlation coefficient is statistically significant in the more experienced driver group may be that *N* is much larger and not because the correlation is stronger. The same interpretation applies to Cohen's results, since *N* differs between inexperienced drivers (1,358) and experienced drivers (103,279).

Another potential caveat is that the group of inexperienced, or young, drivers is more likely to be homogeneous compared with a sample of senior drivers.<sup>33</sup> This potentially biases the correlation test. However, inexperienced drivers will not have an informational advantage over their insurer, as they do not have enough experience of their own driving on which to make inferences on how risky they are. A related study by Arvidsson<sup>34</sup> uses this data set and shows that new policyholders who stayed with the insurer for a year or less are more likely to make a claim than long-term customers. Both groups consist of inexperienced and experienced drivers. The conclusion is that, since insurers do not share claim history, high-risk drivers have an incentive to switch insurer when their type is revealed.

#### *The standard positive correlation test*

Table 5 reports the results from the bivariate probit model of Eqs. (1) and (2) for our sample of new policyholders aged 18 years and over with a vehicle 3–20 years old.

Overall, it seems that the insurance company is able to handle the information asymmetry, since there tends to be no significant correlation in the majority of groups. But,

<sup>33</sup> C&S suggest a note of caution when considering individuals with various driving records and ages. One reason is heteroscedasticity, since the distribution of random shocks will depend on seniority; older individuals are more likely to report a claim due to longer exposure.

<sup>34</sup> Arvidsson (2010).

**Table 5** Correlation test between all risk insurance and culpa

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.140** (0.061)	0.041 (0.046)	0.015 (0.054)	0.034 (0.040)	–0.007 (0.045)	0.032 (0.039)	0.083*** (0.033)	0.020 (0.031)	–0.004 (0.023)	0.074*** (0.017)
95% confidence interval ( $\rho$ )	[0.018, 0.258]	[–0.050, 0.131]	[–0.091, 0.120]	[–0.045, 0.111]	[–0.095, 0.081]	[–0.043, 0.108]	[0.017, 0.148]	[–0.040, 0.080]	[–0.050, 0.042]	[0.040, 0.107]
Log-likelihood	–2,421.026	–4,689.6897	–5,982.340	–9,009.245	–6,960.222	–10,101.646	–15,387.476	–22,013.852	–37,969.87	–59,579.976
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Likelihood ratio test of $\rho=0$ ,	5.083	0.772	0.077	0.703	0.022	0.698	6.155	0.432	0.034	18.748
$\chi^2$ (1)										
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The dependent variables are claims at culpa where the policyholder was held either partially or fully responsible, and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurer's risk classification. The total number of observations of new policyholders with a vehicle 3–20 years old is 295,846 for 2006 and 2007. Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level. See main text for more details.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.



conditional on the risk classification, the correlation coefficient is positively significant at the 5 per cent level for women in the age group 18–21 years, at the 1 per cent level for women in the age group 30–39 years and at the 1 per cent level for policyholders of both sexes in the age group 50+ years.<sup>35</sup> This indicates that there exists residual asymmetric information, which supports the adverse selection/moral hazard prediction.

Women aged 30–39 years and policyholders aged 50+ years may be a result of the chief user problem. Even though we have sorted out claims where some other driver was named in the accident report, there may be cases where the policyholder was falsely named as the car driver. The younger group of women (age group 18–21 years) may be riskier than expected by the insurer; women generally pay premiums that are much lower compared with men in this age group, because the latter are viewed as very risky.

Our results are not consistent with those of Cohen. She did, however, use a sample of inexperienced drivers, and it is unclear if she identified new policyholders based on *contract* or *individual*, which makes a crucial difference: An individual might have stayed with the insurer for a long period of time before deciding to buy a second vehicle. In our data, this second vehicle will appear as a new contract even though the policyholder is loyal. It is therefore important to sort data on *new policyholders*, and perform the analysis on their contracts, rather than sort data on *new contracts* only. Most papers in the literature do not make a clear distinction between policyholder and contract.

The results in Table A4 of the sensitivity analysis of the correlation test include at-fault claims for all drivers; that is, cases where the policyholder *and* other drivers, not included in the contract, caused the accident in the claim. The correlation coefficient also becomes significant for men (age group 30–39 years) and the age group 40–49 years. Table A5, which contains the results for all claims, shows that the correlation coefficient becomes significant for all groups. As the insurer lacks information on additional drivers, we cannot include this information in  $\mathbf{X}$ , and we expect to find a positive correlation in line with our prediction.

Since we use the insurer's risk classification as the control, the result will likely reflect that the market is characterised by asymmetric information. The sensitivity analysis in Appendix A (Tables A1–A6) further shows the importance of an accurate conditioning on the insurers' risk classification and the group to whom we apply the test.

### *Including private information*

Tables 3 and 4 report the marginal effects from estimating the relationship between private information on risky traffic behaviour, more insurance coverage and culpa in Eqs. (3) and (4), respectively. A bivariate probit model is used in groups where there is a significant correlation between Eqs. (3) and (4): Similarly, the equations are estimated independently in groups where there is an insignificant correlation.

Table 6 shows the results from estimating the relationship between private information on risky behaviour and insurance coverage in Eq. (3). The results

---

<sup>35</sup> We also apply Finkelstein and Poterba's (2004) approach to test the correlation between coverage and risk;  $\text{prob}(y=1)=\Phi(\mathbf{X}\beta_1 + c\beta_2)$ , where  $c=\text{ARI}$  with the low deductible. The positive correlation prediction is that  $\beta_2 > 0$ . This test confirms the results from the positive correlation test.

**Table 6** Relationship between new policyholders' private information and extensive coverage

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
Gender	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.253** (0.113)	0.007 (0.007)	0.054*** (0.019)	–0.007 (0.007)	0.061*** (0.020)	0.026*** (0.009)	0.069** (0.032)	0.012* (0.007)	–0.013** (0.006)	–0.025* (0.015)
Other traffic offences	–0.198 (0.133)	–0.008 (0.006)	0.011 (0.022)	0.002 (0.007)	–0.012 (0.025)	–0.008 (0.009)	–0.004 (0.043)	–0.016** (0.008)	–0.018*** (0.007)	–0.092** (0.020)
One conviction: 1973–2007	—	–0.050 (0.019)	–0.119 (0.098)	0.017 (0.019)	0.132 (0.086)	0.004 (0.017)	–0.081 (0.067)	–0.008 (0.009)	–0.023*** (0.007)	0.029* (0.017)
Several convictions: 1973–2007	0.631*** (0.220)	0.014* (0.009)	–0.033 (0.061)	–0.057*** (0.010)	–0.216*** (0.065)	–0.052*** (0.016)	–0.395*** (0.124)	–0.094*** (0.011)	–0.080*** (0.009)	–0.232*** (0.026)
Correlation coefficient ( $\rho$ )	0.143** (0.062)						0.085*** (0.033)			0.077*** (0.017)
95% confidence interval ( $\rho$ )	[0.021, 0.261]						[0.019, 0.149]		[0.021, 0.261]	
Likelihood ratio test of $\rho=0$ , $\chi^2$ (1)	5.240						6.440		20.234	
Log-likelihood	–2,409.5562	–3,403.367	–5,413.4944	–7,769.567	–6,248.2716	–9,082.3895	–15,375.416	–20,391.904	–35,210.911	–59,505.089
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: The reported coefficients are marginal effects from probit estimation of (3): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has all-risk insurance with low deductible or not. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder has one or more on-the-spot fines for speeding, one or more on-the-spot fines for traffic offences, one conviction or more than one conviction for traffic safety violations. The total number of observations is 295,846. (—) indicates that the variable is omitted due to empty cells (The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

**Table 7** Relationship between new policyholders' private information and at-fault claims

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.261 (0.169)	0.007* (0.005)	0.003 (0.004)	0.001 (0.002)	0.004 (0.004)	0.000 (0.002)	−0.091 (0.097)	0.002 (0.001)	0.001 (0.001)	0.070* (0.042)
Other traffic offences	0.350** (0.156)	0.005 (0.004)	0.011*** (0.006)	0.001 (0.002)	0.001 (0.012)	0.002 (0.002)	0.127 (0.107)	0.003*** (0.001)	0.004*** (0.001)	0.078 (0.053)
One conviction: 1973–2007	—	—	—	−0.000 (0.005)	0.012 (0.023)	0.010*** (0.005)	0.299** (0.147)	−0.000 (0.001)	−0.001 (0.001)	0.107*** (0.046)
Several convictions: 1973–2007	0.299 (0.333)	0.003 (0.005)	0.001 (0.010)	0.005 (0.004)	0.008 (0.020)	0.005 (0.004)	0.312 (0.237)	0.001 (0.002)	0.001 (0.001)	0.235*** (0.061)
Correlation coefficient ( $\rho$ )	0.143*** (0.062)						0.085*** (0.033)			0.077*** (0.044)
95% confidence interval ( $\rho$ )	[0.021, 0.261]						[0.019, 0.149]		[0.044, 0.110]	
Likelihood ratio test of $\rho=0$ ,	5.240						6.440		20.334	
$\chi^2$ (1)										
Log-likelihood	−2,409.5562	−1,279.9611	−560.1342	−1,225.6187	−700.1885	−1,005.2104	−15,375.416	−1,577.5631	−2,702.6105	−59,505.089
Prob > $\chi^2$	0.0000	0.0310	0.0122	0.1480	0.6994	0.0349	0.0000	0.0023	0.0003	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (4): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether or not the policyholder was fully, partially or slightly to blame for causing the accidents in the claim. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder has one or more on-the-spot fines for speeding, one or more on-the-spot fines for traffic offences, one conviction or more than one conviction for traffic safety violations. The total number of observations of new policyholders for 2006 and 2007 with a vehicle 3–20 years old is 295,846. (—) indicates that the variable is omitted due to empty or small cells. (As in Table 6, the indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations. In the age groups 18–21 years (men) and 22–25 years (women), the indicator variable for one conviction perfectly predicts cases where the dependent variable is 0. There were 16 and 17 cases, respectively, where this indicator variable was equal to 1, which implies that those with one conviction did not have a reported at-fault claim. The indicator for one conviction variable is therefore omitted due to empty cells in these groups.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

indicate that speeding increases the probability of more insurance, except for the mixed gender and age groups 40–49 and 50+ years. Moreover, private information on other traffic offences and several convictions for traffic safety violations tends to essentially decrease the probability of more insurance coverage. A sensitivity analysis is performed in Appendix C, where we (i) include all private information variables as one single variable and (ii) drop each variable as a crude check for multicollinearity, and the result seems robust. One possible explanation of these results is that speeding is a socially acceptable violation and is not perceived as a risky behaviour, hence the policyholder may not use this information in the decision on whether or not to purchase insurance. Other traffic offences and major infractions are generally viewed as unacceptable, and individuals committing those may perceive themselves as less risky or represent a type that may not bother to purchase insurance. Since we lack information on how the policyholders perceive their risk, it is difficult to fully understand why speeding reflects a varying pattern. Future research could benefit from survey data on infractions, especially speeding, perceived risk and insurance.

Table 7 reports the results from estimating the relationship between private information and at-fault claims in Eq. (4). The results, where significant, indicate that private information on risky traffic behaviour tends to increase the probability of claims where the policyholder is fully or partially at fault, that is, risky drivers have more accidents. A note of caution is that the rather low number of significant variables may be associated with an under-reporting of culpa claims. That is, high-risk drivers who do not purchase extensive insurance report fewer claims to the insurance company.

A potential caveat is that we cannot observe all contracts until they expire since data are censored for contracts that start in 2008 and end in 2009. We therefore perform a sensitivity analysis of the effect of private information on culpa, where we include only new policyholders for 2007 (see Tables B1–B3). The reason is that the censoring may lead to an under-reporting of culpa claims. The results indicate the same pattern as for new policyholders in 2007 and 2008, the conclusion being that our results are not sensitive to the censoring.

We have also performed a sensitivity analysis, where deregistered vehicles are excluded from the analysis. Having the vehicle deregistered may affect the choice of coverage; if the vehicle is not in use, there may be no reason to purchase full coverage. Speeding and one conviction, which is significant at the 10 per cent level in Table 3, do not have a significant effect on extensive coverage in the age group 50+ years, otherwise the results are robust (see Table B4). For this reason, we have also performed a sensitivity analysis of at-fault claims; speeding becomes significant at the 10 per cent level for women aged 18–21 years, while speeding becomes insignificant for men of the same age group (see Table B5). Hence, our results seem robust.

## Conclusions

All in all, our results suggest that there is residual private information in three groups and that high-risk drivers are less likely to purchase full insurance. The data enable us to compare the outcome of C&S with and without private information by using the

Finkelstein and McGarry<sup>2</sup> approach. Taken together, the results in Tables 6 and 7 show that the correlation coefficient is not affected when including private information. If the correlation arises from policyholders' private information on traffic offences, it is reasonable to expect that this would be captured by the correlation test and that the correlation would vanish. However, since high-risk drivers are less likely to purchase full coverage, the correlation may not follow our expectation. In most studies, a potential caveat with the correlation test, no matter the accuracy of conditioning on the insurers' information set, is that the results are biased by information observed by the insurer and not the researcher. This potential caveat is not an issue in this study since we have access to the insurer's risk classification. It may be hazardous, though, to study asymmetric information based on only a coverage correlation test. Even with access to necessary information from the insurer, we can only conclude whether the market is characterised by asymmetric information or not. More importantly, the conclusion may be vicious if the correlation structure does not correspond to our theoretical expectations. The advantage of the Finkelstein and McGarry test is that private information is included explicitly and it is possible to directly observe the outcome of asymmetric information.

It is reasonable to question whether we should expect to find any evidence of information asymmetries in the insurance market. The reason is that, an accurate conditioning on the insurer's risk classification would eliminate any correlation, at least if the risk classification used by the insurer is efficient. We believe that the answer to this question depends on potential information asymmetries in each market, keeping in mind that private information in some markets may be public in others. A general challenge for any empirical analysis of insurance data is the difference in structure across insurance markets. For instance, market heterogeneity, as imposed by laws and regulations, may explain why some markets tend to have a negative correlation, while others tend to have a positive, or even no correlation between risk and coverage. We therefore suggest that empirical work in this area should not try to find a correlation that generally holds for all insurance markets. It is reasonable to believe that the ambiguity found across insurance markets does not necessarily imply a contradiction; it may rather be a consequence of market heterogeneity. We suggest that future research should consider specific market characteristics and subsets of policyholders that are likely to be affected by, or take advantage of, information asymmetries. If high-risk drivers are less likely to purchase full coverage, the market may not be characterised by the positive correlation expected by theory. The solution is to include policyholders' private information in the analysis when studying information asymmetries.

## **Acknowledgements**

I thank Länsförsäkringar AB for insurance data; Lage Niemann and Björn Johansson at Länsförsäkringar AB for helpful discussions about the data, the company and its market. I also thank the Swedish Police and the Swedish National Council of Crime Prevention for data on traffic violations. Thanks to Jan-Eric Nilsson, Lars Hultkrantz, Daniela Andrén, Henrik Andersson and participants at seminars at Örebro University and VTI as well as participants at the World Congress of Risk and Insurance Economics, especially Robert Kreamslehner, for useful comments on previous drafts of this paper. I also thank two anonymous referees for valuable input and thorough comments.

## References

- Åberg, L. (1993) 'Drinking and driving: Intentions, attitudes and social norms of Swedish male drivers', *Accident Analysis & Prevention* 25(3): 289–296.
- Åberg, L. (1998) 'Traffic rules and traffic safety', *Safety Science* 29(3): 205–215.
- Åberg, L. and Rimmö, P.A. (1998) 'Dimensions of aberrant driver behaviour', *Ergonomics* 41(1): 39–56.
- Arvidsson, S. (2010) 'Essays on asymmetric information in the automobile insurance market', Doctoral Dissertation, Örebro Studies in Economics, Vol. 20.
- Bolton, P. and Dewatripont, M. (2005) *Contract Theory*, Cambridge, MA: MIT Press.
- Chiappori, P.-A. and Salanié, B. (1997) 'Empirical contract theory: The case of insurance data', *European Economic Review* 41(3–5): 943–950.
- Chiappori, P.-A. and Salanié, B. (2000) 'Testing for asymmetric information in insurance markets', *Journal of Political Economy* 108(1): 56–78.
- Chiappori, P.A. and Salanié, B. (2003) 'Testing contract theory: A survey of some recent work', in M. Dewatripont, L. Hansen and S. Turnovsky (eds) *Advances in Economics and Econometrics*, Cambridge: Cambridge University Press, Vol. 1, pp. 115–149.
- Cohen, A. (2005) 'Asymmetric information and learning: Evidence from the automobile insurance market', *The Review of Economics and Statistics* 87(2): 197–207.
- Cohen, A. and Siegelman, P. (2010) 'Testing for adverse selection in insurance markets', *The Journal of Risk and Insurance* 77(1): 39–84.
- Cutler, D. ([2000] 2002) 'Health care and the public sector', in A. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Amsterdam: North Holland, Vol. 4, pp. 2143–2243.
- Dahlby, B. (1983) 'Adverse selection and statistical discrimination: An analysis of Canadian automobile insurance', *Journal of Public Economics* 20(1): 121–130.
- Dahlby, B. (1992) 'Testing for asymmetric information in Canadian automobile insurance', in G. Dionne (ed.) *Contributions to Insurance Economics*, Boston, MA: Kluwer Academic, pp. 423–443.
- De Donder, P. and Hindriks, J. (2009) 'Adverse selection, moral hazard and propitious selection', *Journal of Risk and Uncertainty* 38(1): 73–86.
- De Meza, D. and Webb, D. (2001) 'Advantageous selection in insurance markets', *The RAND Journal of Economics* 32(2): 249–262.
- Dionne, G., Gouriéroux, C. and Vanasse, C. (2001) 'Testing for evidence of adverse selection in the automobile insurance market: A comment', *Journal of Political Economy* 109(2): 444–453.
- Fang, H., Keane, M. and Silverman, D. (2008) 'Sources of advantageous selection: Evidence from the Medigap insurance market', *Journal of Political Economy* 116(2): 303–350.
- Finkelstein, A. and McGarry, K. (2006) 'Multiple dimensions of private information: Evidence from the long-term car insurance market', *American Economic Review* 96(4): 938–958.
- Finkelstein, A. and Poterba, J. (2004) 'Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market', *Journal of Political Economy* 112(1): 183–208.
- Forward, S.E. (2006) 'The intention to commit driving violations—A qualitative study', *Transportation Research: Part F: Traffic Psychology and Behaviour* 9(6): 412–426.
- Forward, S.E. (2008) 'Driving violations investigating forms of irrational rationality', Digital comprehensive summaries of Uppsala dissertations from the Faculty of Social Sciences, No. 44, Uppsala University.
- Forward, S.E., Kós-Dienes, D. and Obrenovic, S. (2000) *Invandrare i trafiken, en attitydsundersökning i Värmland och Skaraborgs län* [Immigrants in Traffic, a Study of Attitudes in Two Counties in Sweden], VTI Report 454.
- Guppy, A. (1993) 'Subjective probability of accident and apprehension in relation to self-other bias, age, and reported behavior', *Accident Analysis & Prevention* 25(4): 375–382.
- Hemenway, D. (1990) 'Propitious selection', *The Quarterly Journal of Economics* 105(4): 1063–1069.
- Koufopoulos, K. (2007) 'On the positive correlation property in competitive insurance markets', *Journal of Mathematical Economics* 43(5): 597–605.
- Puelz, R. and Snow, A. (1994) 'Evidence on adverse selection: Equilibrium, signaling and cross-subsidization in the insurance market', *Journal of Political Economy* 102(2): 236–257.
- Rothschild, M. and Stiglitz, J.E. (1976) 'Equilibrium in competitive insurance markets: An essay on the economics of imperfect information', *The Quarterly Journal of Economics* 90(4): 629–649.
- Salanié, B. (2005) *The Economics of Contracts: A Primer*, Cambridge, MA: MIT Press [Originally: *Théorie des Contrats* (1994)].

## **Appendix A**

### **Sensitivity analysis of vehicle age, claims and omitted control variables**

We first apply a sensitivity analysis to the vehicle age restriction, since full coverage may not be motivated for older vehicles due to the value of the car. We apply the positive correlation test for vehicle age 3–15, 3–10 and 3–5 (see Tables A1–A3). The correlation is insignificant for all groups when the vehicle is 3–5 years old, but becomes significant for women aged 30–39 years and the mixed gender group aged 50+ years when the vehicle is 3–10 and 3–15 years old. Hence, the correlation structure does not differ a lot when considering different age intervals for the vehicle, and our findings regarding vehicles 3–20 years old seem to be robust.

We expect the significance level of the correlation coefficient to increase if we consider all at-fault claims. That is, we also include cases where a driver other than the owner was at fault in the accident. As previously mentioned, the insurers do not include additional drivers in their risk classification. When including additional drivers, the correlation coefficient also becomes significant for men in the age group 30–39 years and the mixed gender group aged 40–49 years (see Table A4). When we include all reported claims, the results suggest a significant positive correlation for all groups (see Table A5).

When excluding control variables, the correlation coefficients increase compared to Table A5. This is expected according to the omitted variable bias.

**Table A1** Correlation test between all-risk insurance and culpa, vehicles 3–5 years old

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.307 (0.185)	0.176 (0.198)	-0.023 (0.147)	-0.126 (0.102)	-0.155 (0.107)	-0.037 (0.079)	0.015 (0.068)	0.067 (0.063)	0.005 (0.046)	0.040 (0.034)
Log-likelihood	-249.5207	-366.689	-1,103.2989	-1,661.5886	-1,575.6054	-2,562.3717	-3,933.0821	-6,264.1549	-8,482.2055	-15,204.56
Prob > $\chi^2$	0.0157	0.0033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	330	560	1,623	2,311	2,280	3,523	6,063	8,960	12,530	22,707

Notes: Dependent variables are at-fault claims and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurers' risk classification. The total number of observations is 60,887. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table A2** Correlation test between all-risk insurance and culpa, vehicles 3–10 years old

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.092 (0.090)	-0.012 (0.076)	0.064 (0.064)	0.065 (0.051)	-0.019 (0.055)	0.007 (0.006)	0.102*** (0.041)	0.027 (0.038)	0.042 (0.030)	0.093*** (0.022)
Log-likelihood	-1,266.2341	-1,929.264	-4,020.8968	-5,588.6211	-4,880.3022	-6,989.9253	-10,694.881	-15,189.234	-23,228.236	-37,817.804
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	1,803	2,925	5,625	8,133	6,700	9,661	15,504	21,408	33,488	55,059

Notes: Dependent variables are at-fault claims and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurers' risk classification. The total number of observations is 160,306. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.



**Table A3** Correlation test between all-risk insurance and culpa, vehicles 3–15 years old

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.078 (0.071)	0.054 (0.050)	0.030 (0.056)	0.045 (0.042)	−0.013 (0.047)	0.027 (0.040)	0.089*** (0.035)	0.028 (0.032)	−0.000 (0.025)	0.077*** (0.018)
Log-likelihood	−2,031.650	−3,680.4919	−5,519.0509	−7,942.3165	−6,509.2132	−9,281.796	−14,254.525	−20,201.478	−33,586.787	−51,923.579
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	3,759	7,089	8,384	13,758	9,333	14,435	21,114	30,502	50,174	75,687

Notes: Dependent variables are at-fault claims and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurers' risk classification. The total number of observations is 234,235. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table A4** Correlation test between all-risk insurance and at-fault claims with all drivers

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.010* (0.058)	0.052 (0.043)	0.051 (0.047)	0.044 (0.035)	0.018 (0.038)	0.035 (0.033)	0.115*** (0.028)	0.046** (0.024)	0.058*** (0.017)	0.072*** (0.014)
Log-likelihood	−2,520.1394	−4,880.1598	−6,186.7039	−9,382.2815	−7,296.1497	−10,567.78	−15,954.572	−22,954.946	−40,651.638	−62,459.062
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: Dependent variables are: All at-fault claims (where the policyholder and drivers other than the owner fully, partially or slightly caused the accident in the claim) and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurers' risk classification. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table A5** Correlation test between all-risk insurance and all claims

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
Gender	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.160*** (0.037)	0.158*** (0.027)	0.105*** (0.024)	0.113*** (0.019)	0.099*** (0.022)	0.140*** (0.018)	0.110*** (0.015)	0.133*** (0.012)	0.126*** (0.009)	0.124*** (0.008)
Log-likelihood	-3,437.3003	-6,648.4529	-8,418.0474	-12,931.521	-9,715.8665	-14,361.022	-21,382.128	-30,939.512	-54,158.634	-81,461.713
Prob $> \chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: Dependent variables are all reported claims and all-risk insurance with the low deductible (the highest possible insurance coverage) on insurers' risk classification. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table A6** Correlation test between all-risk insurance and all claims with fewer control variables

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
Gender	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.214*** (0.034)	0.222** (0.025)	0.177*** (0.022)	0.203** (0.018)	0.149*** (0.020)	0.219*** (0.016)	0.166*** (0.014)	0.202*** (0.012)	0.191*** (0.009)	0.164*** (0.007)
Log-likelihood	-3,920.1135	-7,277.2245	-9,543.3883	-14,417.56	-10,824.715	-16,003.139	-24,080.2	-34,334.118	-60,780.65	-89,182.44
Prob $> \chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: Dependent variables are all reported claims and all-risk insurance with the low deductible (the highest possible insurance coverage) on insurers' risk classification. Excluded variables are vehicle risk classification and vehicle age. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

## Appendix B

### Sensitivity analysis of new policyholders for 2007 and registered cars

To investigate whether the results are sensitive to the censoring for 2008, we perform a sensitivity analysis of the correlation test on new policyholders for 2007; that is, contracts where we can observe the whole lifespan. The results indicate that there exists a positive correlation between risk and coverage for women aged 30–39 years and the mixed gender age group 50+ years. The conclusion is that our results regarding new policyholders for 2007 and 2008 do not suffer from a serious under-reporting of claims due to the censoring of outcomes of some contracts signed in 2008 (see Tables B1–B3).

**Table B1** Correlation test between all-risk insurance and at-fault claims for new policyholders 2007

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Correlation coefficient ( $\rho$ )	0.002 (0.085)	0.025 (0.054)	–0.028 (0.066)	0.035 (0.049)	0.002 (0.056)	0.051 (0.045)	0.110*** (0.040)	0.000 (0.037)	–0.004 (0.028)	0.061*** (0.020)
Log-likelihood	–1,500.1075	–3,095.4631	–3,782.6156	–5,736.3686	–4,405.199	–6,570.0699	–9,712.517	–14,089.007	–24,400.295	–37,368.932
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	3,489	7,753	6,420	11,736	6,933	11,273	15,542	23,379	40,200	57,987

*Notes:* Dependent variables are at-fault claims and all-risk insurance with the low deductible (the highest possible insurance coverage). Independent variables correspond to the insurers' risk classification. Standard errors are in parentheses.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table B2** Relationship between new policyholders' (2007) private information and at-fault claims

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Mixed group	Females	Males	Mixed group	Mixed group
Speeding	0.019* (0.014)	0.005 (0.006)	0.007 (0.005)	0.001 (0.003)	0.009* (0.006)	0.002 (0.003)	−0.139 (0.120)	0.002 (0.002)	0.001 (0.001)	0.073 (0.050)
Other traffic offences	0.025** (0.017)	0.008* (0.005)	0.006 (0.006)	0.003 (0.003)	0.009 (0.008)	0.005* (0.003)	0.162 (0.124)	0.005*** (0.002)	0.005*** (0.002)	0.112* (0.050)
One conviction: 1973–2007	—	—	—	0.003 (0.007)	—	0.006 (0.006)	0.278* (0.172)	−0.000 (0.001)	−0.001 (0.002)	0.142*** (0.053)
Several convictions: 1973–2007	0.038 (0.040)	0.002 (0.007)	0.008 (0.019)	0.007 (0.006)	—	0.012** (0.007)	0.522** (0.260)	−0.001 (0.002)	0.001 (0.001)	0.294*** (0.071)
Correlation coefficient ( $\rho$ )							0.114*** (0.040)			0.065*** (0.021)
Log-likelihood	−345.27846	−931.24069	−387.46899	−4,884.304	−444.22095	−745.42265	−9,698.8307	−1,110.0727	−1,945.9379	−37,317.798
Prob > $\chi^2$	0.0102	0.0186	0.00215	0.1004	0.4312	0.0352	0.0000	0.0015	0.0000	0.0000
N	3,489	7,753	6,420	11,736	6,993	11,273	15,542	23,379	40,200	57,987

Notes: The reported coefficients are marginal effects from a probit estimation of Eq. (4) for new policyholders in 2007: A bivariate probit estimation of Eqs. (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations. (—) indicates that the variable was omitted due to empty or small cells. Standard errors are in parentheses.

\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table B3** Relationship between new policyholders' (2007) private information and coverage

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.060*** (0.030)	0.005 (0.009)	0.103*** (0.025)	−0.014 (0.009)	0.086*** (0.025)	0.031*** (0.004)	0.036 (0.040)	0.008* (0.009)	−0.034*** (0.009)	−0.026 (0.019)
Other traffic offences	−0.043** (0.018)	−0.010 (0.007)	0.026 (0.028)	0.001 (0.009)	−0.024 (0.031)	−0.022 (0.012)	0.024 (0.053)	−0.013 (0.010)	−0.016** (0.008)	−0.080*** (0.025)
One conviction: 1973–2007	—	−0.054 (0.019)	−0.204* (0.090)	0.014 (0.022)	0.345*** (0.098)	0.019 (0.023)	−0.120*** (0.082)	−0.012 (0.011)	−0.028*** (0.009)	−0.030 (0.021)
Several convictions: 1973–2007	0.273*** (0.099)	0.035*** (0.013)	0.056 (0.086)	−0.065*** (0.014)	−0.206* (0.097)	−0.035*** (0.022)	−0.567*** (0.172)	−0.115*** (0.014)	−0.080*** (0.012)	−0.229*** (0.033)
Correlation coefficient ( $\rho$ )							0.114*** (0.041)			0.065*** (0.021)
Log-likelihood	1,139.8335	−2,155.7906	−3,381.2977	−4,884.304	−3,945.5089	−5,813.1039	−9,698.8307	−12,940.977	−22,408.904	−37,317.798
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	3,489	7,753	6,420	11,736	6,933	11,273	15,542	23,379	40,200	57,987

Notes: The reported coefficients are marginal effects from a probit estimation of Eq. (3) for new policyholders in 2007 only: A bivariate probit estimation of (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding and one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations. (—) indicates that the variable was omitted due to empty cells. Standard errors are in parentheses.

\*\*\*, \*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

**Table B4** Relationship between private information and coverage for registered cars only

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.280*** (0.115)	0.009 (0.008)	0.054*** (0.019)	−0.007 (0.007)	0.060*** (0.020)	0.028*** (0.009)	0.071** (0.032)	0.014** (0.007)	−0.012** (0.006)	−0.021 (0.015)
Other traffic offences	−0.198 (0.134)	−0.007 (0.006)	0.013 (0.022)	0.005 (0.007)	−0.007 (0.026)	−0.006 (0.010)	−0.010 (0.043)	−0.017** (0.008)	−0.016** (0.008)	−0.100*** (0.020)
One conviction: 1973–2007	—	−0.054 (0.018)	−0.133 (0.095)	0.022 (0.020)	0.141 (0.090)	0.006 (0.018)	−0.083 (0.067)	−0.010 (0.009)	−0.024*** (0.007)	−0.027 (0.017)
Several convictions: 1973–2007	0.620*** (0.222)	0.014* (0.009)	−0.042 (0.061)	−0.059*** (0.011)	−0.226*** (0.065)	−0.051*** (0.017)	−0.384*** (0.126)	−0.096*** (0.011)	−0.081*** (0.009)	−0.239*** (0.026)
Correlation coefficient ( $\rho$ )	0.135** (0.063)						0.080** (0.033)			0.066*** (0.017)
Log-likelihood	−2,378.7886	−3,325.6375	−5,337.9121	−7,630.7391	−6,167.2002	−8,959.5147	−15,149.563	−20,065.285	−34,395.144	−58,051.223
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,503	12,093	10,167	18,354	10,797	17,599	24,238	36,355	61,948	90,473

*Notes:* The reported coefficients are marginal effects from a probit estimation of Eq. (3) for registered vehicles only: A bivariate probit estimation of (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding and one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations. (—) indicates that the variable was omitted due to empty cells. Standard errors are in parentheses.

\*\*\*, \*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

**Table B5** Relationship between private information and at-fault claims for registered cars only

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.278* (0.170)	0.007 (0.005)	0.003 (0.004)	0.000 (0.002)	0.004 (0.005)	0.000 (0.002)	−0.090 (0.097)	0.002 (0.001)	0.001 (0.001)	0.072* (0.042)
Other traffic offences	0.356** (0.157)	0.006 (0.004)	0.011*** (0.006)	0.001 (0.001)	0.001 (0.012)	0.003 (0.002)	0.128 (0.107)	0.003*** (0.002)	0.004*** (0.001)	0.077 (0.054)
One conviction: 1973–2007	—	—	—	−0.000 (0.005)	0.013 (0.024)	0.011*** (0.005)	0.301** (0.147)	−0.000 (0.001)	−0.001 (0.001)	0.236*** (0.061)
Several convictions: 1973–2007	0.292 (0.334)	0.003 (0.005)	0.001 (0.010)	0.005 (0.004)	0.008 (0.020)	0.006 (0.005)	0.324 (0.239)	0.001 (0.002)	0.001 (0.001)	0.109*** (0.046)
Correlation coefficient ( $\rho$ )	0.135** (0.063)						0.080* (0.033)			0.066*** (0.017)
Log-likelihood	2,378.7886	−1,262.9085	−558.4515	−1,213.741	−697.99861	−1,000.7921	−15,149.563	−1,571.4454	−2,689.8275	−58,051.223
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,503	12,093	10,167	18,354	10,797	17,599	24,238	36,355	61,948	90,473

*Notes:* The reported coefficients are marginal effects from a probit estimation of Eq. (4) for registered vehicles only: A bivariate probit estimation of (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding and one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations.. (—) indicates that the variable was omitted due to empty cells. Standard errors are in parentheses.

\*\*\*, \*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

## Appendix C

### Sensitivity analysis with private information as one variable and check for multicollinearity

**Table C1** Relationship between private information and coverage, with private information as one variable

<i>Age (years)</i>	<i>18–21</i>		<i>22–25</i>		<i>26–29</i>		<i>30–39</i>		<i>40–49</i>	<i>50+</i>
<i>Gender</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Mixed group</i>	<i>Mixed group</i>
Private information	0.134 (0.085)	0.004 (0.005)	0.032 (0.015)	−0.0104 (0.005)	0.027* (0.016)	0.001 (0.007)	0.016 (0.025)	0.018*** (0.005)	−0.037*** (0.005)	−0.075*** (0.011)
Correlation coefficient ( $\rho$ )	0.142** (0.063)						0.082** (0.003)			0.076*** (0.017)
Log-likelihood	−2,416.4068	−3,406.493	−5,416.451	−7,780.6151	−6,256.962	−9,091.5317	−15,385.189	−20,426.229	−35,229.284	−59,548.019
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<i>N</i>	5,669	12,689	10,373	19,026	10,989	18,048	24,238	370,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from a probit estimation of Eq. (3). A bivariate probit estimation of (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by a dummy variable taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences, one conviction or more than one conviction for traffic safety violations. Standard errors are in parentheses.

\*\*\* and \* represent significance at 1%, 5% and 10% levels, respectively.



**Table C2** Relationship between private information and at-fault claims, with private information as one variable

<i>Age (years)</i>	<i>18–21</i>		<i>22–25</i>		<i>26–29</i>		<i>30–39</i>		<i>40–49</i>	<i>50+</i>
<i>Gender</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Females</i>	<i>Males</i>	<i>Mixed group</i>	<i>Mixed group</i>
Private information	0.329*** (0.118)	0.006 (0.003)	0.008 (0.004)	0.002 (0.002)	0.004 (0.004)	0.003 (0.002)	0.101 (0.067)	0.002 (0.001)	0.001 (0.001)	0.128*** (0.031)
Correlation coefficient ( $\rho$ )	0.142** (0.063)						0.082** (0.033)			0.076*** (0.017)
Log-likelihood	-2,416.4068	-1,262.9085	-5,560.221	-1,226.3031	-700.651	-1,008.1871	-15,386.189	-1,579.6755	-2,707.2364	-59,548.019
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<i>N</i>	5,669	12,689	10,373	19,026	10,989	18,048	24,238	36,355	63,912	93,204

*Notes:* The reported coefficients are marginal effects from a probit estimation of Eq. (4): A bivariate probit estimation of (3) and (4) is used where there is a significant correlation between the residuals. Private information for all groups is represented by a dummy variable taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences, one conviction or more than one conviction for traffic safety violations. Standard errors are in parentheses.

\*\*\*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

**Table C3** Relationship between new policyholders' private information and extensive coverage (the variable "one conviction" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.253** (0.113)	0.007 (0.007)	0.054*** (0.019)	−0.007 (0.007)	0.061*** (0.020)	0.026*** (0.009)	0.068** (0.032)	0.011* (0.007)	−0.013** (0.006)	−0.026* (0.015)
Other traffic offences	−0.198 (0.133)	−0.008 (0.006)	0.010 (0.022)	0.002 (0.007)	−0.011 (0.025)	−0.007 (0.009)	−0.006 (0.043)	−0.017** (0.008)	−0.019*** (0.007)	−0.093** (0.020)
One conviction: 1973–2007	—	—	—	—	—	—	—	—	—	—
Several convictions: 1973–2007	0.631*** (0.220)	0.014* (0.009)	−0.033 (0.061)	−0.057*** (0.010)	−0.217*** (0.065)	−0.052*** (0.016)	−0.393*** (0.124)	−0.093*** (0.011)	−0.077*** (0.009)	−0.232*** (0.026)
Correlation coefficient ( $\rho$ )	0.143** (0.062)						0.085*** (0.033)			0.077*** (0.017)
95% confidence interval ( $\rho$ )	[0.021, 0.261]						[0.019, 0.149]		[0.043, 0.110]	
Likelihood ratio test of $\rho=0$ , $\chi^2$ (1)	5.240						6.397		20.234	
Log-likelihood	−2,409.5562	−3,404.344	−5,414.0683	−7,770.0259	−6,249.4701	−9,082.4134	−15,378.011	−20,392.353	−35,216.119	−59,509.262
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (3): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has all-risk insurance with low deductible or not. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences and more than one conviction for traffic safety violations. The total number of observations is 295,846. (—) indicates that the variable is omitted. Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\*\*, \*\* and \* represent significance at 1%, 5% and 10% levels, respectively.

**Table C4** Relationship between new policyholders' private information and at-fault claims (the variable "one conviction" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.261 (0.169)	0.007* (0.005)	0.003 (0.004)	0.001 (0.002)	0.004 (0.004)	0.000 (0.002)	−0.086 (0.097)	0.002 (0.001)	0.001 (0.001)	0.074* (0.042)
Other traffic offences	0.350** (0.156)	0.005 (0.004)	0.011*** (0.006)	0.001 (0.002)	0.001 (0.012)	0.002 (0.002)	0.136 (0.107)	0.003*** (0.001)	0.003*** (0.001)	0.083 (0.053)
One conviction: 1973–2007	—	—	—	—	—	—	—	—	—	—
Several convictions: 1973–2007	0.299 (0.333)	0.003 (0.005)	0.001 (0.010)	0.005 (0.004)	0.008 (0.020)	0.005 (0.004)	0.299 (0.237)	0.001 (0.002)	0.001 (0.001)	0.220*** (0.060)
Correlation coefficient ( $\rho$ )	0.143*** (0.062)						0.085*** (0.033)		0.077*** (0.017)	
95% confidence interval ( $\rho$ )	[0.021, 0.261]						[0.019, 0.149]		[0.043, 0.110]	
Likelihood ratio test of $\rho=0$ , $\chi^2(1)$	5.240						6.397		20.132	
Log-likelihood	−2,409.5562	−1,279.8218	−560.082	−1,225.619	−700.408	−1,008.1716	−15,378.011	−1,577.5796	−2,703.1858	−59,509.262
Prob > $\chi^2$	0.0000	0.0427	0.0186	0.1095	0.6631	0.1217	0.0000	0.0014	0.0003	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (4): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has reported a claim where s/he is partially or fully at fault. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences and more than one conviction for traffic safety violations. The total number of observations of new policyholders for 2006 and 2007 with a vehicle 3–20 years old is 295,846. (—) indicates that the variable was omitted. Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level. \*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table C5** Relationship between new policyholders' private information and extensive coverage (the variable "several convictions" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.265** (0.112)	0.007 (0.007)	0.054*** (0.019)	–0.007 (0.007)	0.062*** (0.020)	0.026*** (0.009)	0.069** (0.032)	0.010 (0.007)	–0.015** (0.006)	–0.032** (0.015)
Other traffic offences	–0.186 (0.132)	–0.007 (0.006)	0.009 (0.021)	0.001 (0.007)	–0.016 (0.025)	–0.010 (0.009)	–0.012 (0.043)	–0.024** (0.008)	–0.025*** (0.007)	–0.109** (0.020)
One conviction: 1973–2007	—	–0.050 (0.019)	–0.119 (0.098)	0.021 (0.019)	0.134 (0.086)	0.007 (0.017)	–0.078 (0.067)	–0.001 (0.009)	–0.015** (0.007)	0.016 (0.017)
Several convictions: 1973–2007	—	—	—	—	—	—	—	—	—	—
Correlation coefficient ( $\rho$ )	0.145** (0.063)						0.084*** (0.034)		0.075*** (0.017)	
95% confidence interval ( $\rho$ )	[0.023, 0.262]						[0.019, 0.150]		[0.042, 0.109]	
Likelihood ratio test of $\rho=0$ , $\chi^2(1)$	5.373						6.350		20.234	
Log-likelihood	–2,413.6827	–3,404.757	–5,413.6373	–7,781.2479	–6,252.0242	–9,087.0585	–15,381.522	–20,426.096	–35,248.327	–59,553.855
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: The reported coefficients are marginal effects from probit estimation of (3): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has all-risk insurance with low deductible or not. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences and one conviction for traffic safety violations. The total number of observations is 295,846. (—) indicates that the variable is omitted. (The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table C6** Relationship between new policyholders’ private information and at-fault claims (the variable “several convictions” is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.262 (0.169)	0.007* (0.005)	0.003 (0.004)	0.001 (0.002)	0.004 (0.004)	0.000 (0.002)	−0.087 (0.097)	0.002 (0.001)	0.001 (0.001)	0.070* (0.042)
Other traffic offences	0.357** (0.155)	0.005 (0.004)	0.011*** (0.006)	0.001 (0.002)	0.001 (0.012)	0.003 (0.002)	0.134 (0.107)	0.003*** (0.001)	0.004*** (0.001)	0.099* (0.052)
One conviction: 1973–2007	—	—	—	−0.000 (0.005)	0.012 (0.023)	0.010*** (0.005)	0.294** (0.147)	−0.000 (0.001)	−0.001 (0.001)	0.090** (0.045)
Several convictions: 1973–2007	—	—	—	—	—	—	—	—	—	—
Correlation coefficient ( $\rho$ )	0.145** (0.063)						0.084*** (0.034)			0.075*** (0.017)
95% confidence interval ( $\rho$ )	[0.023,0.262]						[0.019,0.150]			[0.042,0.109]
Likelihood ratio test of $\rho=0, \chi^2(1)$	5.373						6.350			20.234
Log-likelihood	−2,413.6827	−1,279.7437	−559.9683	−1,226.5634	−700.30713	−1,006.1805	−15,381.522	−1,577.6553	−2,702.6825	−59,553.855
Prob > $\chi^2$	0.0000	0.0333	0.0122	0.1744	0.6470	0.0414	0.0000	0.0015	0.0002	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (4): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has reported a claim where s/he is partially or fully at fault. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one or several on-the-spot fines for traffic offences and one conviction for traffic safety violations. The total number of observations of new policyholders for 2006 and 2007 with a vehicle 3–20 years old is 295,846. (—) indicates that the variable was omitted. (The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations. In the age groups 18–21 years (men) and 22–25 years (women), the indicator variable for one conviction perfectly predicts cases where the dependent variable is 0. There were 16 and 17 cases, respectively, where this indicator variable was equal to 1, which implies that those with one conviction did not have a reported at-fault claim. The indicator for one conviction variable is therefore omitted due to empty cells in these groups.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table C7** Relationship between new policyholders' private information and extensive coverage (the variable "other traffic offences" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.237** (0.113)	0.007 (0.007)	0.055*** (0.019)	–0.007 (0.007)	0.060*** (0.019)	0.025*** (0.009)	0.069** (0.032)	0.010 (0.007)	–0.015** (0.006)	–0.033** (0.015)
Other traffic offences	—	—	—	—	—	—	—	—	—	—
One conviction 1973–2007	—	–0.050 (0.019)	–0.119 (0.098)	0.018 (0.019)	0.132 (0.086)	0.003 (0.017)	–0.081 (0.067)	–0.010 (0.009)	–0.023*** (0.007)	0.032* (0.017)
Several convictions: 1973–2007	0.620*** (0.220)	0.012 (0.009)	–0.031 (0.061)	–0.057*** (0.010)	–0.218*** (0.065)	–0.053*** (0.016)	–0.395*** (0.124)	–0.097*** (0.011)	–0.082*** (0.009)	–0.243*** (0.026)
Correlation coefficient ( $\rho$ )	0.139** (0.061)						0.084*** (0.033)			0.077*** (0.017)
95% confidence interval ( $\rho$ )	[0.017, 0.257]						[0.019, 0.1501]		[0.043, 0.1098]	
Likelihood ratio test of $\rho=0$ , $\chi^2(1)$	4.971						6.440		20.0373	
Log-likelihood	–2,413.0993	–3,404.339	–5,413.6167	–7,769.5986	–6,248.3902	–9,082.7117	–15,376.085	–2,039.039	–35,213.885	–59,517.011
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (3): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has all-risk insurance with low deductible or not. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one conviction and more than one conviction for traffic safety violations. The total number of observations is 295,846. (—) indicates that the variable is omitted. (The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level. \*\*\*\*\*, \*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table C8** Relationship between new policyholders' private information and at-fault claims (the variable "other traffic offences" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	0.303*	0.009**	0.004	0.001	0.004	0.000	–0.085	0.002	0.001	0.077*
	(0.166)	(0.005)	(0.004)	(0.002)	(0.004)	(0.002)	(0.097)	(0.001)	(0.001)	(0.042)
Other traffic offences	—	—	—	—	—	—	—	—	—	—
One conviction: 1973–2007	—	—	—	–0.000	0.012	0.011***	0.307**	–0.000	–0.001	0.109***
				(0.005)	(0.023)	(0.005)	(0.146)	(0.001)	(0.001)	(0.046)
Several convictions: 1973–2007	0.342	0.003	0.004	0.005	0.008	0.006	0.329	0.001	0.001	0.245***
	(0.326)	(0.005)	(0.013)	(0.004)	(0.020)	(0.005)	(0.237)	(0.002)	(0.001)	(0.060)
Correlation coefficient ( $\rho$ )	0.139**						0.084***			0.077***
	(0.061)						(0.033)			(0.017)
95% confidence interval ( $\rho$ )	[0.017, 0.257]						[0.019, 0.1501]		[0.043, 0.1098]	
Likelihood ratio test of $\rho=0$ , $\chi^2(1)$	4.971						6.440		20.0373	
Log-likelihood	–2,413.0993	–1,280.5727	–562.73012	–1,225.6841	–700.19299	–1,005.9401	–15,376.085	–1,580.5595	–2,706.7356	–59,517.011
Prob $> \chi^2$	0.0000	0.0544	0.0122	0.1132	0.6288	0.0360	0.0000	0.0023	0.0003	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: The reported coefficients are marginal effects from probit estimation of (4): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has reported a claim where s/he is partially or fully at fault. Private information for all groups is represented by four dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for speeding, one conviction and more than one conviction for traffic safety violations. The total number of observations of new policyholders for 2006 and 2007 with a vehicle 3–20 years old is 295,846. (—) indicates that the variable was omitted. ( The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations. In the age groups 18–21 years (men) and 22–25 years (women), the indicator variable for one conviction perfectly predicts cases where the dependent variable is 0. There were 16 and 17 cases, respectively, where this indicator variable was equal to 1, which implies that those with one conviction did not have a reported at-fault claim. The indicator for one conviction variable is therefore omitted due to empty cells in these groups.) Standard errors are in parenthesis. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

**Table C9** Relationship between new policyholders' private information and extensive coverage (the variable "speeding" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	—	—	—	—	—	—	—	—	—	—
Other traffic offences	-0.172 (0.133)	-0.007 (0.006)	0.018 (0.022)	0.003 (0.007)	-0.005 (0.025)	-0.0083 (0.009)	-0.002 (0.043)	-0.014* (0.008)	-0.020*** (0.007)	-0.096*** (0.020)
One conviction 1973–2007	—	-0.050 (0.019)	-0.118 (0.097)	0.017 (0.019)	0.132 (0.086)	0.006 (0.017)	-0.076 (0.067)	-0.008 (0.009)	-0.023*** (0.007)	0.029* (0.017)
Several convictions: 1973–2007	0.648*** (0.220)	0.014* (0.009)	-0.038 (0.061)	-0.057*** (0.010)	-0.230*** (0.065)	-0.052*** (0.016)	-0.392*** (0.124)	-0.097*** (0.011)	-0.081*** (0.009)	-0.234*** (0.026)
Correlation coefficient ( $\rho$ )	0.144** (0.063)						0.085*** (0.033)			0.077*** (0.017)
95% confidence interval ( $\rho$ )	[0.020, 0.261]						[0.019, 0.150]		[0.043, 0.110]	
Likelihood ratio test of $\rho=0$ , $\chi^2$ (1)	5.214						6.358		20.2515	
Log-likelihood	-2,412.9904	-3,403.865	-5,417.6812	-7,770.1189	-6,253.2154	-9,086.6262	-15,378.287	-20,393.317	-35,213.232	-59,507.862
Prob > $\chi^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

Notes: The reported coefficients are marginal effects from probit estimation of (3): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has all-risk insurance with low deductible or not. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations. The total number of observations is 295,846. (—) indicates that the variable is omitted. (The indicator variable for one conviction is always 0 for women in the age group 18–21years, which implies that no one in this group had one conviction for traffic safety violations.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.



**Table C10** Relationship between new policyholders' private information and at-fault claims (the variable "speeding" is dropped)

Age (years)	18–21		22–25		26–29		30–39		40–49	50+
	Females	Males	Females	Males	Females	Males	Females	Males	Mixed group	Mixed group
Speeding	—	—	—	—	—	—	—	—	—	—
Other traffic offences	0.377** (0.154)	0.006 (0.004)	0.011*** (0.006)	0.001 (0.002)	0.001 (0.012)	0.002 (0.002)	0.121 (0.107)	0.003*** (0.001)	0.004*** (0.001)	0.088* (0.053)
One conviction: 1973–2007	—	—	—	–0.000 (0.005)	0.012 (0.023)	0.010*** (0.005)	0.294** (0.146)	–0.000 (0.001)	–0.001 (0.001)	0.110*** (0.046)
Several convictions: 1973–2007	0.303 (0.333)	0.003 (0.005)	0.001 (0.010)	0.005 (0.004)	0.008 (0.020)	0.005 (0.004)	0.304 (0.238)	0.001 (0.002)	0.001 (0.001)	0.241*** (0.061)
Correlation coefficient ( $\rho$ )	0.144** (0.063)						0.085*** (0.033)			0.077*** (0.017)
95% confidence interval ( $\rho$ )	[0.020, 0.261]						[0.019, 0.150]		[0.043, 0.110]	
Likelihood ratio test of $\rho=0$ , $\chi^2(1)$	5.214						6.358		20.2515	
Log-likelihood	–2,412.9904	–1,280.888	–560.349	–1,225.6762	–700.77538	–1,005.2193	–15,378.287	–1,578.4797	–2,703.2382	–59,507.862
Prob $> \chi^2$	0.0000	0.0651	0.0156	0.1127	0.7206	0.0334	0.0000	0.0023	0.0026	0.0000
N	5,669	12,689	10,373	19,026	10,989	18,048	24,686	37,250	63,912	93,204

*Notes:* The reported coefficients are marginal effects from probit estimation of (4): A bivariate probit model is used where there is a significant correlation between the residuals. The dependent variable is an indicator of whether the policyholder has reported a claim where s/he is partially or fully at fault. Private information for all groups is represented by dummy variables taking the value 1 if the policyholder had one or several on-the-spot fines for traffic offences, one conviction and more than one conviction for traffic safety violations. The total number of observations of new policyholders for 2006 and 2007 with a vehicle 3–20 years old is 295,846. (—) indicates that the variable was omitted. (The indicator variable for one conviction is always 0 for women in the age group 18–21 years, which implies that no one in this group had one conviction for traffic safety violations. In the age groups 18–21 years (men) and 22–25 years (women), the indicator variable for one conviction perfectly predicts cases where the dependent variable is 0. There were 16 and 17 cases, respectively, where this indicator variable was equal to 1, which implies that those with one conviction did not have a reported at-fault claim. The indicator for one conviction variable is therefore omitted due to empty cells in these groups.) Standard errors are in parentheses. The 95% confidence interval of  $\rho$  is in brackets. The critical value of the  $\chi^2$  distribution with 1 df is 3.841 at the 95% confidence level.

\*\*\*\* and \* indicate significance at 1%, 5% and 10% levels, respectively.

## Appendix D

1. Demographic characteristics of the policyholder: Individual id-number, year of birth, gender, home district and self-reported number of kilometres driven per year.
2. Residential area risk classification: The actuarial predicted risk in the neighbourhood where the policyholder lives. Each type of insurance coverage (traffic insurance, limited damage insurance and all-risk insurance) has a classification. All policyholders have each classification regardless of coverage.
3. Car characteristics: Vehicle model, make, construction year, size of engine and vehicle-id.
4. Vehicle risk classification: The actuarial risk classification regarding the vehicle. As with residential area risk classification, each type of insurance coverage has a risk classification regarding the vehicle.
5. Private information: The number of on-the-spot fines for speeding or other traffic offences of the policyholder in the period 2004–2007, and the number of convictions a policyholder had from 1973 to 2007.
6. The type of policy purchased: Traffic insurance (required if the car is in use, but not if it is deregistered), limited damage insurance, all-risk insurance (not generally required for new cars since most manufacturers provide insurance) and Additional insurance.
7. Deductible choice: The only contract providing deductible choice (high or low deductible) is all-risk insurance.
8. Premium: The price of the insurance policy.
9. Period covered: From date and to date for each period in the contracts. The number of days with insurance is 1–365 days during one period.
10. Realisation of risk: Claims submitted by the policyholder and information on which insurance covers the claim. It is also possible to identify the level of at-fault in the claim (none, partial or fully responsible).
11. Driver information: The insurer's information on the identity of the reported driver in an accident (not necessarily the policyholder), age, gender and personal identity number and private information according to (5). Note that additional drivers are private information to the policyholder since the premium is not dependent on drivers other than the vehicle owner.
12. Other variables: Household identity, two or more policyholders in the same household share the same household-id.

## About the Author

**Sara Forsstedt** has a PhD in Economics and is currently a researcher at the Swedish National Road and Transport Research Institute (VTI). Her research is concerned with the consequences of asymmetric information of risky behaviour in the automobile insurance market.