Econometrics of Insurance with Multidimensional Types

Gaurab Aryal
Boston University

Isabelle Perrigne
Rice University

Quang Vuong
New York University

Haiqing Xu
University of Texas Austin

September 2024

This paper draws from Aryal, Perrigne and Vuong (2016). It extends it in important directions while including a new section on estimation. We thank Pierre-André Chiappori, Liran Einav, Matt Shum and Ken Wolpin as well as two referees and the Editor for constructive comments. We also benefited from participants' comments at the Stanford Institute for Theoretical Economics, North American Meeting of the Econometric Society and seminars at the Australian National University, Collegio Carlo Alberto, Columbia University, Georgetown University, London School of Economics, National University of Singapore, Paris School of Economics, Sciences Politiques-Paris, Stanford University, University of Fortaleza, University of Pennsylvania, University of Sydney and University of Wisconsin at Madison. Isabelle Perrigne and Quang Vuong gratefully acknowledge financial support from the National Science Foundation through grant SES 1148149.

Correspondence to Isabelle Perrigne: iperrigne@gmail.com

Abstract

In this paper, we address the identification and estimation of insurance models where insurees have private information about their risk and risk aversion. The model includes random damages and allows for several claims, while insurees choose from a finite number of coverages. We show that the joint distribution of risk and risk aversion is nonparametrically identified despite bunching due to multidimensional types and a finite number of coverages. Our identification strategy exploits the observed number of claims as well as an exclusion restriction, and a full support assumption. Furthermore, our results apply to any form of competition. We propose a novel estimation procedure combining nonparametric estimators and GMM estimation that we illustrate in a Monte Carlo study.

Keywords: Insurance, Identification, Nonparametric Estimation, Multidimensional Adverse Selection, Risk Aversion.

Econometrics of Insurance with Multidimensional Types

G. Aryal, I. Perrigne, Q. Vuong & H. Xu

1 Introduction

Insurance is a long-studied topic in economics and is at the core of recent empirical research. Rothschild and Stiglitz (1976) and Stiglitz (1977) provide benchmark models of insurance under private information on insurees' risk. In empirical studies, testing adverse selection in risk has generated a large number of papers with mixed results.[1] The empirical literature shows that insurance markets involve heterogeneity in both risk and risk aversion. See, e.g., Finkelstein and McGarry (2006) for long-term care insurance, Cohen and Einav (2007) for automobile insurance, and Fang, Keane and Silverman (2008) for health insurance.[2] As noted in these papers, heterogeneity in risk aversion may overturn the prediction of the benchmark adverse selection model that risk and insurance coverage have a positive correlation. For instance, a low-risk individual may buy higher coverage because of high risk aversion and conversely. Whether risk aversion or risk is the primary determinant of the demand for insurance has distinct welfare implications and policy recommendations. Thus, a model of insurance needs to incorporate insurees' heterogeneity in risk aversion resulting in multidimensional screening and pooling.

Letting each insuree be characterized by two parameters capturing his/her risk preference and risk, this paper addresses the nonparametric identification and estima-

[1]See Chiappori and Salanié (2000) for the most well known test and Cohen and Siegelman (2010) for a survey of empirical findings.

[2]See also Cutler, Finkelstein and McGarry (2008) and Einav and Finkelstein (2011) for surveys.

tion of the joint distribution of risk and risk aversion from a finite number of coverage choices with random damages and multiple claims. Allowing for a flexible dependence between risk and risk aversion is important for policy recommendations. Moreover, our identification result requires minimal assumptions on the supply side. In particular, it does not rely either on an insurer's model of coverage offering or on how insurers compete, thereby avoiding well-known complexities of optimal contracting with multi-dimensional types. See e.g. Rochet and Chone (1998) and Rochet and Stole (2003). Identification is a key step for the econometric and empirical analysis of structural models. First, it highlights which variations in the data allow one to identify model primitives. Second, our identification argument is constructive. It provides the basis for our proposed estimation method.

We consider a finite number of automobile coverages of the form 'premium and deductible' although our results apply to other insurance markets and/or other types of coverages as discussed later. Since there is no one-to-one mapping between the deductible and the insuree's private information due to multidimensional types (risk and risk aversion) and a finite number of coverages, the number of claims plays a key role in identifying the marginal distribution of risk. To identify the joint distribution of risk and risk aversion, we exploit an exclusion restriction and a support assumption that requires sufficient variations in some exogenous characteristics.

The previous results are derived with two offered coverages of the form 'premium and deductible' under the widely used specifications of a Constant Absolute Risk Aversion (CARA) utility function and a Poisson distribution for the number of accidents. The CARA parameter and the mean of the Poisson distribution then measure the risk preference and risk for each insuree, respectively. We show that our identification results extend to more than two offered coverages and a larger class of models beyond the CARA attitude toward risk and the Poisson distribution for the number of accidents. Observing more coverages helps identify the joint distribution of risk and risk aversion by alleviating the full support assumption. Moreover, our results also extend to health insurance coverages of the form 'premium, deductible, and copayment' with a fixed or proportional deductible. Regarding estimation, we develop a new multi-step procedure that is computationally friendly. The procedure combines several nonparametric estimators (kernel and sieves estimators) and GMM estimation. A Monte Carlo study illustrates our estimation procedure.

Our paper differs on several aspects from the previous literature on the identification and estimation of models under incomplete information. Multidimensional adverse selection leads to bunching or pooling, making model identification a challenging problem. In particular, identification can no longer rely on the one-to-one equilibrium mapping(s) between the agent's unobserved continuous types and his observed outcome(s)/action(s).[3] A finite number of contracts often leads to pooling with similar identification issues.[4] We develop a different identification strategy relying on the insurees' choice of coverage and the observed number of claims. Given that we do not rely on the optimality of the offered coverages, our results apply to any form of competition in the insurance industry. This result contrasts with the previous literature and provides a novel perspective on the identification of models under incomplete information. In particular, the estimation of the model primitives can no longer rely on inversion as in Guerre, Perrigne and Vuong (2000) or quantiles as in Marmer and Shneyerov (2012) or Luo, Perrigne and Vuong (2018). As a consequence, our paper proposes a new estimation procedure.

The paper is organized as follows. Section 2 presents the model, whereas Section 3 studies its identification and an important extension with more than two contracts. Section 4 presents our new estimation procedure and a Monte Carlo study. Section 5 concludes. An online appendix collects some proofs, auxiliary results and several extensions. See Aryal, Perrigne, Vuong and Xu (2024).

[3]See Luo, Perrigne and Vuong (2017) who study identification of nonlinear pricing models with multiple types relying on Armstrong (1996) model and Aryal and Zincenko (2023) who study identification of Rochet and Chone (1998) model. In contrast, Kong, Perrigne and Vuong (2022) exploit the bidders' multiattribute bids in auctions to avoid the complexity of the optimal mechanism.

[4]Crawford and Shum (2007) and Gayle and Miller (2015) circumvent this issue by considering as many contracts as agents' (one dimensional) discrete types.

2 A Model of Insurance

This section presents a model in which insurees have private information about their risk and risk aversion when buying insurance from a finite number of coverages. In the presence of multivariate private information and/or a finite number of coverages, pooling arises as individuals of different types choose the same coverage. To fix ideas and in the spirit of the early literature on adverse selection, we consider automobile insurance throughout the paper although our framework also applies to (say) homeowner and rental insurance. See the Appendix for a discussion of an application to health insurance.

MOTIVATION

Our model draws from Stiglitz (1977), wherein insurees are heterogeneous in their probability of accident (risk), which is their private information, but are homogeneous in risk aversion. However, Finkelstein and McGarry (2006) and Cohen and Einav (2007) find that heterogeneity in risk aversion might be more important than heterogeneity in risk across insurees. Thus, we consider that risk aversion is as heterogeneous and private like the probability of an accident. Consequently, asymmetric information becomes bidimensional. In addition, Stiglitz (1977) assumes that there can be at most one accident with fixed damage. In reality, there might be more than one accident during the policy period, and ex-ante, every accident involves random damage.

Ignoring this bidimensional feature may affect insurance policy design. For instance, an insuree with a low probability of an accident and a high risk aversion may buy a contract with high coverage, i.e., a low deductible. Similarly, an insuree with a high probability of an accident and a low risk aversion may buy a contract with low coverage, i.e., a high deductible. These two examples contrast with Stiglitz (1977) predictions as the former should choose a low coverage and the latter a high coverage when insurees have homogeneous preferences. Furthermore, when risk and risk aversion are negatively correlated, this leads to advantageous selection where high coverages are bought by insurees with low risk but high risk aversion. Using a probit regression for the choice of deductible and a Poisson regression for the number of claims on a set of insurees' characteristics, Cohen and Einav (2007) show, for instance, that married, educated and female insurees tend to have fewer accidents while buying a high coverage. Their structural analysis confirms that these insurees tend to be more risk averse. It is, therefore, crucial to consider heterogeneity in risk aversion. Similarly, the distribution of damages may have an important effect on the insuree's choice of coverage. For instance, at a given level of risk and risk aversion, a higher expected damage will induce the insuree to choose more coverage.

In view of this, our model includes multiple accidents with random damages and heterogeneity in risk and risk aversion. Although insurance contracts may take various forms, we consider the benchmark case of premium-deductible contracts. Our results, however, extend to other forms of insurance coverage. For instance the Appendix contains an extension to health insurance with a copayment per claim in addition to a premium and a deductible for the coverage period.

Model Assumptions

We make the following assumptions. In our model, the insuree's risk θ is the expected number of accidents during the coverage period whereas the parameter a measures the insuree's risk aversion. They satisfy the following assumptions.

Assumption A1:

(i) An insuree's utility function exhibits Constant Absolute Risk Aversion (CARA), i.e., $U(x;a) = -\exp(-ax)$ *where* $a > 0$,

(ii) The types (θ, a) *are jointly distributed as* $F(\cdot,\cdot)$ *with positive density* $f(\cdot,\cdot)$ *on its support* $\Theta \times \mathcal{A} = (\underline{\theta}, \overline{\theta}) \times (\underline{a}, \overline{a}) \subset I\!R_{++} \times I\!R_{++}$,

(iii) Conditional on θ, *the number* J *of accidents each insuree may have follows a Poisson distribution* $\mathcal{P}(\theta)$, *i.e.,* $p_j(\theta) = \Pr[J = j|\theta] = e^{-\theta}\theta^j/j!$ *for* $j = 0, 1, \ldots,$

(iv) Each accident involves a damage D_j *independently distributed as* $H(\cdot)$ *on* $(0, \overline{d}) \subset I\!R_+$ *for* $j = 1, \ldots, J$.

Individual and car characteristics are introduced later in Assumption A2. By A1-(i), the utility function is increasing and concave. The CARA specification has two main advantages: (i) it keeps the model tractable, and (ii) the attitude toward risk is independent of initial wealth. In most empirical settings, the agent's wealth is not observed, making the CARA utility specification quite convenient. These properties have made the CARA utility popular in the theoretical and empirical literature. By A1-(ii), each insuree is characterized by a pair (θ, a), which is his/her private information. Assumption A1-(iii) specifies the distribution of accidents as Poisson with mean θ. This distribution is widely used in actuarial science to model the number of accidents for a given individual. The distribution $F(\theta)$ of risk is left unspecified, so the distribution of the number of accidents in the population is a flexible nonparametric mixture of Poisson distributions, namely $\Pr(J = j) = \int_\Theta p_j(\theta) dF(\theta)$.[5]

The combination of the CARA utility and the Poisson distribution is convenient to model the expected utility and the associated certainty equivalent. Relaxing the CARA and/or Poisson specifications is possible at the cost of obtaining implicit expressions for the expected utility. We consider alternative specifications for the agent's utility function and the distribution of accidents in the online appendix. Assumption A1-(iii,iv) imply that the number J of accidents and corresponding damages are independent of a and (θ, a), respectively. These independence assumptions will be relaxed in Section 3 by introducing the insurees' characteristics, such as their driving experience, which then allows for unconditional dependence between the number J of accidents with a and damages D_j with (θ, a).

The model primitives are the joint distribution of risk and risk aversion and the damage distribution, i.e., $[F(\cdot,\cdot), H(\cdot)]$. Each insuree is characterized by a pair of types (θ, a) from $F(\cdot,\cdot)$, which is left unspecified. The insuree chooses among a finite menu of insurance contracts of the form $[t, dd]$, where dd is the deductible per accident. The insuree chooses the contract that maximizes his expected utility and pays the corresponding premium t. In case of an accident with damage below the deductible, the insuree pays for it. Otherwise, the insurer pays the damage above the deductible, and the insuree pays the deductible.

Insuree's Choice of Coverages

Let C be the number of available contracts. To simplify the presentation, we con-

[5]Using a mixture of Poisson distributions to model the number of accidents in a population dates back to Greenwood and Yule (1920) where the mixing distribution is a Gamma distribution thereby leading to a Negative Binomial distribution for the population. Cohen and Einav (2007) consider a log normal mixture of Poisson distributions.

sider $C = 2$. Extending the analysis to $C > 2$ is presented in Section 3. Let (t_1, dd_1) and (t_2, dd_2) denote two coverages with $0 < t_1 < t_2$ and $\overline{d} > dd_1 > dd_2 \geq 0$ so that no contract dominates the other. This ordering is the only requirement we make on the observed coverages. It is related to rational offering and choice. Otherwise, we would have $t_1 < t_2$ and $dd_1 \leq dd_2$ making contract 1 the natural choice for insurees. These coverage terms do not need to satisfy profit maximizing conditions for the insurer(s) allowing us to be agnostic about the market structure of the insurance industry. We note that our setting can include full coverage, which corresponds to $dd_2 = 0$. Moreover, the highest deductible dd_1 should be smaller than the maximum damage $\overline{d}$ to rationalize buying some insurance. Indeed, the no insurance option, which corresponds to $(t_0, dd_0) = (0, \overline{d})$, would dominate (t_1, dd_1) if $dd_1 = \overline{d}$ since $t_1 > 0$.

For a (θ, a)-individual with wealth w, his expected utility with coverage (t, dd) for $t \geq 0$ and $0 \leq dd \leq \overline{d}$ is $V(t, dd; \theta, a, w) \equiv \mathrm{E}\left[U\left(w - t - \sum_{j=0}^{J} \min\{dd, D_j\}; a\right) |\theta\right]$ where $D_0 \equiv 0$ by convention. Under A1, we obtain

$$\begin{aligned}
V(t, dd; \theta, a, w) &= p_0(\theta)U(w - t; a) + p_1(\theta)\mathrm{E}[U(w - t - \min\{dd, D_1\}; a)] \\
&\quad + p_2(\theta)\mathrm{E}[U(w - t - \min\{dd, D_1\} - \min\{dd, D_2\}; a)] + \dots \\
&= -p_0(\theta)e^{-a(w-t)} - p_1(\theta)e^{-a(w-t)}\mathrm{E}[e^{a\min\{dd,D_1\}}] \\
&\quad - p_2(\theta)e^{-a(w-t)}\mathrm{E}[e^{a\min\{dd,D_1\}}]\mathrm{E}[e^{a\min\{dd,D_2\}}] - \dots \\
&= -e^{-a(w-t)}\left[p_0(\theta) + p_1(\theta)\phi_a(dd) + p_2(\theta)\phi_a^2(dd) + \dots\right] \\
&= -e^{-a(w-t)}e^{-\theta}\left(1 + \frac{\theta\phi_a(dd)}{1!} + \frac{\theta^2\phi_a^2(dd)}{2!} + \dots\right) \\
&= -e^{-a(w-t)+\theta[\phi_a(dd)-1]}, \qquad (1)
\end{aligned}$$

with $\phi_a(dd) \equiv \mathrm{E}[e^{a\min\{dd,D\}}] < \infty$ where the expectation is with respect to the random damage D. In particular, $\phi_a(dd) \geq 1$ with equality only if $dd = 0$, as $a > 0$. The expression $\phi_a(dd)$ can be interpreted as the expected loss in utils of an accident with deductible dd for an individual with risk aversion a. The first equality in (1) considers all the possible number of accidents and their respective costs for a (θ, a)-individual buying insurance (t, dd). The second equality uses the CARA utility function and A1-(i,iv). The third equality uses that damages are identically distributed by A1-(iv). Lastly, the fourth equality relies on the Poisson distribution of accidents by A1-(iii).

The (θ, a)-individual chooses the contract that maximizes his expected utility or equivalently his certainty equivalent. The certainty equivalent $CE(t, dd; \theta, a, w)$ of insurance coverage is defined as the amount of certain wealth for the insuree that will give him/her the same level of utility when he/she has coverage, i.e., $-\exp(-aCE(t, dd; \theta, a, w)) = V(t, dd; \theta, a, w)$. Thus, by (1) we have for $t \geq 0$ and $0 \leq dd \leq \overline{d}$

$$CE(t, dd; \theta, a, w) = w - t - \frac{\theta[\phi_a(dd) - 1]}{a}. \tag{2}$$

When comparing different insurance coverages for a (θ, a)-individual, i.e., their certainty equivalents, the individual wealth w then cancels out, which makes the choice of the CARA and Poisson specifications quite convenient. Therefore, wealth need not be observed. The next lemma establishes the monotonicity in (θ, a) of the certainty equivalent as well as the frontier that partitions the $\Theta \times \mathcal{A}$ space into two subsets $\mathcal{C}_1$ and $\mathcal{C}_2$ of individuals choosing coverages 1 and 2, respectively.

Lemma 1: *Let A1 hold.*
(i) When $dd = 0$ (full coverage), the certainty equivalent (2) reduces to $w - t$. When $dd > 0$, the certainty equivalent (2) decreases in both risk and risk aversion.
(ii) The frontier separating $\mathcal{C}_1$ and $\mathcal{C}_2$ is given by

$$\theta(a) = \frac{a(t_2 - t_1)}{\phi_a(dd_1) - \phi_a(dd_2)} = \frac{t_2 - t_1}{\int_{dd_2}^{dd_1} e^{aD}[1 - H(D)]dD}, \tag{3}$$

which is decreasing in a. Every (θ, a)-individual below (resp. above) this frontier prefers coverage 1 to coverage 2 (resp. 2 to 1).[6]

The proof is given in the Appendix. The first part of (i) is expected as wealth is reduced only by the premium with full coverage. The second part of (i) is also intuitive as the certainty equivalent, and the utility move together. Regarding (ii), the frontier between $\mathcal{C}_1$ and $\mathcal{C}_2$ is the locus of (θ, a)-insurees who are indifferent between the two contracts, i.e., for whom $CE(t_1, dd_1; \theta, a, w) = CE(t_2, dd_2; \theta, a, w)$. This frontier is independent of wealth w. The denominator of (3) is the difference in expected utility losses from an accident between the two coverages for an individual with risk aversion a.

[6]As a matter of fact, Lemma 1 also applies when $(t_1, d_1) = (0, \overline{d})$ (no insurance), and $(t_2, dd_2) = (t, dd)$ with $t > 0$ and $0 \leq dd < \overline{d}$.

Figure 1 illustrates the choice between two coverages with a Uniform damage distribution on $[0, 10^4]$. In agreement with Cohen and Einav (2007), the range of the CARA parameter a is $[10^{-4}, 10^{-3}]$. The range of the parameter θ is $[0.1, 1]$. The two coverages are $(t_1, dd_1) = (600, 1000)$ and $(t_2, dd_2) = (850, 500)$. The bold curve represents the frontier between the two coverages following (3) with individuals above it preferring coverage 2 to coverage 1. Considering a point on this frontier (say) $(\theta, a) = (0.371, 0.005)$, Figure 1 also displays the certainty equivalent isocurves for coverages 1 and 2 in dashed and dotted curves, respectively. These certainty equivalents decrease as θ or a increases.[7] For completeness, Figure 1 displays a bold dashed curve in the southwest corner corresponding to the frontier between no insurance and coverage 1, with individuals above it preferring coverage 1. This frontier is obtained from (3), by letting the no insurance corresponds to a zero premium and a deductible at $\overline{d} = 10^4$, i.e., to $(t_0, dd_0) = (0, 10^4)$, where $\overline{d}$ represents the car value.

We remark that there is no exclusion since all individuals are willing to buy insurance because coverage 1 always dominates no insurance, i.e., $CE(t_1, dd_1; \theta, a) \geq CE(0, \overline{d}; \theta, a)$ for all (θ, a) since the individual rationality constraints are always satisfied. Figure 1 can be interpreted as an illustration that the insurance industry must serve everyone, i.e., it is subject to a universal service requirement. In contrast, if the premium for coverage 1 goes above 619, the bold dashed curve will shift upward above the $(\underline{a}, \underline{\theta})$ point so that individuals below this curve would prefer no coverage and are therefore excluded. If insurance is mandatory, individuals below the corresponding bold curve (frontier between the two coverages) are forced to buy insurance and will choose coverage 1.

3 Identification

In this section, we study the identification of the joint distribution of risk and risk aversion $F(\theta, a)$ and the damage distribution $H(\cdot)$. Although we do not impose any parameterization on $F(\cdot, \cdot)$ or $H(\cdot)$, our identification analysis is semiparametric since

[7] The top north-east bold curve labelled 'Frontier 2 vs 3' can be omitted at this time and is discussed in Section 3.

the insurees' utility function and their probability of accidents are parameterized by a and θ, respectively. We discuss in the online appendix how to relax the CARA and the Poisson specifications to other parameterizations. Our identification analysis shows the key role played by the number of accidents, an exclusion restriction and a support assumption. The identification problem is to recover uniquely the distributions $F(\cdot,\cdot)$ and $H(\cdot)$ from observables. We observe the contract (t, dd) purchased by each insuree, as well as all his/her J claims with corresponding damages $(D_1, \ldots, D_J)$.[8] Hereafter, insurance is mandatory as when the insurer is subject to universal service. Otherwise, our identification results hold conditional on buying insurance. We first consider the case of two offered coverages and discuss later the benefits of having more coverages.

Introducing Covariates

We introduce some observed variables X characterizing the insuree and his/her car. Variables related to the insuree may contain age, gender, education, marital status, location and driving experience. Variables related to the car may include car mileage, business use, car value, power, model, and make. The structure becomes $[F(\theta, a|X), H(D|X)]$ as such variables may affect the insuree's risk and risk aversion as well as damages. For instance, damages with an expensive car are likely larger than those with an inexpensive one, ceteris paribus. The next assumption specifies the data-generating process. It maintains the CARA and Poisson specifications in A1-(i, iii) while extending A1-(ii, iv) to allow for the characteristics X.

Assumption A2: *The tuplets $(\theta, a, X, J, D_1, \ldots, D_J)$ are i.i.d. across individuals, and*
(i) CARA utility function as in A1-(i),
(ii) $(\theta, a)|X \sim F(\cdot,\cdot|X)$ with positive density $f(\cdot,\cdot|X)$ on its support $\Theta(X)\times\mathcal{A}(X) = (\underline{\theta}(X), \overline{\theta}(X))\times(\underline{a}(X), \overline{a}(X)) \subset I\!R_{++}\times I\!R_{++}$,
(iii) Poisson distribution for the number J of accidents as in A1-(iii),
(iv) The damages $D_j, j = 1, \ldots, J$ are i.i.d. as $H(\cdot|X)$ on $(0, \overline{d}(X)) \subset I\!R_+$.

[8] We abstract away from the truncation issue and assume that all the accidents and damages are observed. In the online appendix, we consider the case when only J^* claims with their corresponding damages $(D_1, \ldots, D_{J^*}$ are observed due to the truncation at the deductible dd. See Aryal, Perrigne, Vuong and Xu (2024).

Assumption A2 parallels A1. As noted, A2-(iii) states that the number of accidents J depends only on the insuree's risk θ through the Poisson distribution, with θ being the expected number of accidents. It implies that J is independent of (a, X) given θ, i.e., $J \perp (X, a)|\theta$. Similarly, A2-(iv) implies that damages are independent of (θ, a) given (J, X), i.e., $(D_1, \ldots, D_J) \perp (\theta, a)|(J, X)$. It should be noted that utilities and number of accidents indirectly depend on the insuree/car characteristics X through a and θ which depend on X by A2-(ii). In particular, A2 allows for *unconditional* correlations between the number J of accidents with insurees's risk aversion a and between damages D_j and insuree's types (θ, a).

The offered coverages may also depend on the vector of characteristics X as $(t_1(X), dd_1(X))$ and $(t_2(X), dd_2(X))$ with $0 < t_1(X) < t_2(X)$ and $\overline{d}(X) > dd_1(X) > dd_2(X) \geq 0$. Following insurance regulations, insurers may not be allowed to use some of the insurees' observed characteristics as discriminatory tools in the coverage terms (t, dd). Thus, (t, dd) may not depend on all the X variables. Hereafter, we let $\mathcal{S}_A$ denote the support of a random vector A and $\mathcal{S}_{A|b}$ the support of A conditional on the value b of a random vector B. Moreover, to simplify, we assume that the upper bounds $\overline{\theta}(x)$, $\overline{a}(x)$, and $\overline{d}(x)$ are finite for every $x \in \mathcal{S}_X$. Hence, all moments of J exist and its moment-generating function is well-defined. Otherwise, our identification arguments hold straightforwardly using characteristic functions.

Hereafter, we show how coverage choices combined with sufficient variations in some exogenous characteristics nonparametrically identify $f(\theta, a|X)$ thereby offering flexibility on the dependence between risk and risk aversion.[9] To begin, the damage distribution $H(\cdot|X)$ is identified on its support $(0, \overline{d}(X))$ using A2-(iv) since all the accidents and damages are observed. It remains to identify the joint distribution $F(\theta, a|X)$.

Identification of $F_{\theta|X}(\cdot|\cdot)$

The first step identifies the conditional distribution of risk $F_{\theta|X}(\cdot|\cdot)$ by exploiting

[9]Using a log-normal joint distribution for (θ, a), Cohen and Einav (2007) find a counterintuitive positive correlation suggesting that the observed contracts are suboptimal, i.e., the insurer could increase his profit by increasing their low deductibles which are more compatible with a negative correlation.

the observed number of accidents/claims J and its nonparametric mixture. Specifically, the probability of J conditional on the characteristics $X = x$ is

$$\Pr[J = j|x] = \int_{\underline{\theta}(x)}^{\overline{\theta}(x)} \Pr[J = j|\theta, x] dF_{\theta|X}(\theta|x) = \int_{\underline{\theta}(x)}^{\overline{\theta}(x)} e^{-\theta} \frac{\theta^j}{j!} dF_{\theta|X}(\theta|x),$$

where the mixing distribution $F_{\theta|X}(\cdot|x)$ is left unspecified.

For insurees with characteristics x, the moment-generating function $M_{J|X}(\cdot|x)$ of the number of claims is

$$\begin{aligned} M_{J|X}(t|x) &= \mathrm{E}[e^{Jt}|X = x] = \mathrm{E}\left\{\mathrm{E}[e^{Jt}|\theta, X]|X = x\right\} \\ &= \mathrm{E}\left\{\mathrm{E}[e^{Jt}|\theta]|X = x\right\} = \mathrm{E}\left\{e^{\theta(e^t-1)}|X = x\right\} \\ &= M_{\theta|X}(e^t - 1|x), \end{aligned}$$

where the third and fourth equalities follow from A2-(iii) and the moment-generating function of the Poisson distribution with parameter θ. In particular, this equation shows that $M_{J|X}(\cdot|x)$ exists on $\mathbb{R}$ because the moment-generating function $M_{\theta|X}(\cdot|x)$ of θ given $X = x$ exists on $\mathbb{R}$ as θ has compact support given $X = x$. Moreover, letting $u = e^t - 1$ gives

$$M_{\theta|X}(u|x) = M_{J|X}(\log(1 + u)|x) = \mathrm{E}[(1 + u)^J|X = x] \tag{4}$$

for all $u \in (-1, +\infty)$. Hence $M_{\theta|X}(\cdot|x)$ is identified in a neighborhood of 0, thereby identifying the density $f_{\theta|X}(\cdot|\cdot)$ of θ given X. See, e.g., Billingsley (1995, p.390).[10] This result exploits the fact that the Poisson distribution belongs to the class of additively closed distributions whose nonparametric mixture is identified. See Rao (1992) and the online appendix.

Identification of $F_{a|\theta,X}[a(\theta, X)|\theta, X]$

The second step considers the probability that a θ-insuree with characteristics X chooses the coverage $(t_1(X), dd_1(X))$. We define a discrete variable χ, which takes

[10] Alternatively, because $M_{\theta|X}(\cdot|x)$ exists in a neighborhood of 0, then all the moments of θ given $X = x$ are identified by the kth derivatives $M^{(k)}_{\theta|X}(0|x) = \mathrm{E}[\theta^k|X = x]$ for $k = 0, 1, \ldots, \infty$. Since θ given x has a bounded support, we are in the class of Hausdorff moment problems, which are always determinate, i.e., the distribution of θ given x is uniquely determined by its moments. For a comprehensive treatment of the moment problem, see Shohat and Tamarkin (1943).

values 1 and 2 depending on whether the insuree chooses the coverage $(t_1(X), dd_1(X))$ or $(t_2(X), dd_2(X))$, i.e., whether his/her pair (θ, a) belongs to the regions $\mathcal{C}_1(X)$ or $\mathcal{C}_2(X)$ of individuals choosing contract 1 or 2, respectively given characteristics X. Thus, from Lemma 1-(ii), $\chi = 1$ is also equivalent to $a \leq a(\theta, X)$ where $a(\cdot, X)$ is the inverse of the frontier $\theta(\cdot, X)$ with

$$\theta(a, X) \quad = \quad \frac{t_2(X) - t_1(X)}{\int_{dd_2(X)}^{dd_1(X)} e^{aD}(1 - H(D|X))dD} \tag{5}$$

from (3). Our identification strategy exploits variations of this frontier in X. In particular, even if the deductible does not vary with X, the premium and/or the damage distribution likely depend on some X.

To ensure that the frontier (5) partitions $\Theta(X) \times \mathcal{A}(X)$ into the two nonempty sets $\mathcal{C}_1(X)$ and $\mathcal{C}_2(X)$, we assume that $\underline{\theta}(X) < \theta(\underline{a}(X), X)$ and $\theta(\overline{a}(X), X) < \overline{\theta}(X)$. In particular, $\mathcal{C}_1(X)$ includes the lowest type individual $(\underline{\theta}(X), \underline{a}(X))$ while $\mathcal{C}_2(X)$ includes the highest type individual $(\overline{\theta}(X), \overline{a}(X))$. For $j = 1, 2$, let $\nu_j(X)$ denote the proportion of insurees with characteristics X choosing the coverage $(t_j(X), dd_j(X))$. Thus, $\nu_j(X) > 0$ for $j = 1, 2$. Such proportions are identified from the data.

The probability $\Pr[\chi = 1|\theta, X = x]$ that a (θ, x)-individual chooses the lowest coverage contract (t_1, dd_1) is

$$F_{a|\theta,X}[a(\theta, x)|\theta, x] = \frac{f_{\theta|\chi,X}(\theta|1, x)\nu_1(x)}{f_{\theta|X}(\theta|x)} \tag{6}$$

by Bayes' rule. Since $f_{\theta|X}(\cdot|\cdot)$ is identified from the first step and $\nu_1(x)$ is identified from the data, it remains to identify $f_{\theta|\chi,X}(\cdot|1, x)$. Applying the same argument as in Step 1, but now conditioning on $\chi = 1$ as well, we obtain

$$\begin{aligned} M_{J|\chi,X}[t|1, x] \quad &= \quad \mathrm{E}[e^{Jt}|\chi{=}1, X{=}x] = \mathrm{E}\{\mathrm{E}[e^{Jt}|\theta, a, X]|\chi{=}1, X{=}x\} \\ &= \quad M_{\theta|\chi,X}[e^t - 1|1, x], \end{aligned}$$

where the second equality follows from the equivalence between conditioning on (θ, a, χ, X) and conditioning on (θ, a, X), while the third equality follows from A2-(iii) as before. Thus, for every $x \in \mathcal{S}_X$, $f_{\theta|\chi,X}(\cdot|1, x)$ is identified by its moment generating function

$$M_{\theta|\chi,X}(u|1, x) = M_{J|\chi,X}(\log(1 + u)|1, x)$$

for all $u \in (-1, +\infty)$. Hence, by (6), $F_{a|\theta,X}[a(\theta,x)|\theta,x]$ is identified for every $\theta \in (\underline{\theta}(x), \overline{\theta}(x))$. That is, we identify the conditional distribution of a given θ on the frontier $a(\theta,x)$ separating the two subsets $\mathcal{C}_1(x)$ and $\mathcal{C}_2(x)$ that partition the set $\Theta(x) \times \mathcal{A}(x)$.

Identification of $F(\theta, a|X)$

For policy counterfactuals the analyst needs to identify $F(\cdot,\cdot|x)$ on the whole support $\Theta(x) \times \mathcal{A}(x)$. This constitutes the third step of identification in which we make an exclusion restriction and a support assumption involving some characteristics Z included in X to achieve identification of the distribution $F_{a|\theta,X}(\cdot|\cdot,\cdot)$ on its support.

We partition the vector of the insuree or car characteristics X into (X_0, Z).

Assumption A3: *We assume that X satisfies the following*
(i) $a \perp Z|(\theta, X_0)$
(ii) $\forall(a,\theta,x_0) \in \mathcal{S}_{a\theta X_0}$*, there exists* $z \in \mathcal{S}_{Z|\theta x_0}$ *such that* $a(\theta,x_0,z) = a$.

Assumption A3-(i) is an exclusion restriction. It requires that some variable Z is independent of risk aversion conditionally on the other variables X_0 (and risk θ). Assumption A3-(ii) is a full support assumption that requires the frontier $a(\theta, X_0, Z)$ to vary sufficiently with Z. In particular, Z needs to be continuous since a is continuous.

In the case of automobile insurance, several variables are potential candidates for Z. For instance, controlling insurees' characteristics such as age and others, Cohen and Einav (2007) empirically find that the engine size and the years of license are not related to risk aversion. Under the CARA specification, the car value, which acts as a proxy for wealth, could also be a good candidate since CARA risk aversion is independent of wealth. Nonetheless, these variables affect the frontier (5) through the premia, the deductibles, and the distribution of damages. Indeed, variations in the frontier arise from variations in premia but also through the difference in expected losses. For instance, a large engine or car value is less likely to lead to damage in the interval $[dd_2, dd_1]$. Thus, a large engine or car value will give a larger value in the denominator of (5) than a low engine or car value.

The combination of an exclusion restriction and a full support assumption is not new in the econometrics literature. See, e.g., Matzkin (2003) and Imbens and Newey (2009). In empirical industrial organization, this includes Berry and Haile (2014) in the nonparametric identification of a demand and supply model for differentiated products. In auctions with selective entry, Gentry and Li (2014) and Chen, Gentry, Li and Lu (2023) assume the existence of a continuous entry cost shifter with full support that affects entry but not the private value distribution. In an application to oil tract lease auctions, Kong (2018) finds that the amount of land offered for auction outside the area is an entry cost shifter satisfying such requirements.

Given A3, for any $(a, \theta, x_0) \in \mathcal{S}_{a\theta X_0}$ we have

$$F_{a|\theta,X_0}(a|\theta, x_0) = F_{a|\theta,X_0}[a(\theta, x_0, z)|\theta, x_0] = F_{a|\theta,X_0,Z}[a(\theta, x_0, z)|\theta, x_0, z],$$

where the first equality uses A3-(ii) and the second equality uses A3-(i). Note that $a(\cdot, \cdot, \cdot)$ is identified from (5) since the premia and deductible are observed while the distribution of damage $H(\cdot|\cdot)$ is identified from claim data. Identification of $F(\theta, a|x_0, z)$ follows from the identification of $F_{a|\theta,X}[a(\theta, x)|\theta, x]$ in Step 2 where $x = (x_0, z)$. This result is formally stated in the next proposition.

Proposition 1: *Suppose two offered coverages and damages are observed for each insuree. Under A2–A3, the structure* $[F(\cdot, \cdot|X), H(\cdot|X)]$ *is identified.*

Despite pooling, due to multidimensional types and a finite number of coverages, Proposition 1 shows that the model primitives are identified by exploiting the number of accidents and sufficient variations in some exogenous variable Z conditionally independent of risk aversion. In particular, our identification argument does not require optimality of the offered coverages. This argument is novel in the identification of models under incomplete information.

In the absence of the full support assumption A3-(ii), the previous argument shows that we can still point identify $F_{\theta|X}(\cdot|\cdot)$ on $\mathcal{S}_{\theta X}$ as well as the conditional distribution $F_{a|\theta,X}(a|\theta, x) = F_{a|\theta,X_0}(a|\theta, x_0)$ on the range of $a(\theta, x_0, \cdot)$ when z varies, i.e. on $\{(a, \theta, x_0) : a = a(\theta, x_0, z), z \in \mathcal{S}_{Z|\theta x_0}, (\theta, x_0) \in \mathcal{S}_{\theta X_0}\}$. See also Section 4. Moreover, assuming that this range is an interval $[a_*(\theta, x_0), a^*(\theta, x_0)]$, where $a_*(\theta, x_0) = \inf_{z \in \mathcal{S}_{Z|\theta x_0}} a(\theta, x_0, z)$ and $a^*(\theta, x_0) = \sup_{z \in \mathcal{S}_{Z|\theta x_0}} a(\theta, x_0, z)$, we can bound the conditional distribution $F(a|\theta, x_0)$ by

$$0 \leq F_{a|\theta,X_0}(a|\theta, x_0) \leq F_{a|\theta,X_0}(a_*(\theta, x_0)|\theta, x_0)$$

$$\text{and} \quad F_{a|\theta,X_0}(a^*(\theta,x_0)|\theta,x_0) \leq F_{a|\theta,X_0}(a|\theta,x_0) \leq 1$$

for $\underline{a}(x_0) \leq a \leq a_*(\theta,x_0)$ and $a^*(\theta,x_0) \leq a \leq \overline{a}(x_0)$, respectively. These bounds are sharp as there is no information on $[\underline{a}(x_0), a_*(\theta,x_0))$ and $(a^*(\theta,x_0), \overline{a}(x_0)]$.[11] It should also be noted that having a larger number of coverages $C > 2$ can only improve the identification results as a larger number of frontiers of the form (5) is more likely to cover the whole support $\Theta(x) \times \mathcal{A}(x_0)$ when Z varies as discussed next.

Beyond Two Coverages

We now consider more than two offered coverages. Let (t_c, dd_c), $c = 1, \ldots, C \geq 2$, be C offered contracts. We omit the insuree/car characteristics to simplify the notations. As before, we require that no observed coverage dominates the others:

$$0 < t_1 < \ldots < t_C \text{ and } \overline{d} > dd_1 > \ldots > dd_C \geq 0. \tag{7}$$

We refer to (7) as the revealed preference (RP) condition, since otherwise some contracts will be irrelevant. It extends the condition that we have for $C = 2$. Moreover, this condition is easily verifiable in the data. Let $\theta_{c,c+1}(a)$ define the frontier or indifference locus between coverages (t_c, dd_c) and (t_{c+1}, dd_{c+1}). It is given by an equation similar to (3). By Lemma 1-(i), each frontier is decreasing and (θ, a)-individuals below (resp. above) the curve $\theta_{c,c+1}(\cdot)$ prefers coverage (t_c, dd_c) over coverage (t_{c+1}, dd_{c+1}) (resp. (t_{c+1}, dd_{c+1}) over (t_c, dd_c)).

The next lemma ensures that the $C-1$ frontiers $\theta_{c,c+1}(\cdot)$ do not cross and lie on top of each other as c increases from 1 to $C-1$.

Lemma 2: *Let A1 hold and the coverages (t_c, dd_c), $c = 1, \ldots, C \geq 2$ satisfy the RP condition (7). The frontiers $\theta_{c,c+1}(\cdot)$ between coverages (t_c, dd_c) and $(t_{c+1},\ dd_{c+1})$ for $c = 1, \ldots, C-1$ satisfy $\theta_{1,2}(\cdot) < \ldots < \theta_{C-1,C}(\cdot)$ on $[\underline{a}, \overline{a}]$ if and only if*

$$\frac{t_{c+2} - t_{c+1}}{t_{c+1} - t_c} > \frac{\int_{dd_{c+2}}^{dd_{c+1}} e^{\underline{a}D}[1 - H(D)]\ dD}{\int_{dd_{c+1}}^{dd_c} e^{\underline{a}D}[1 - H(D)]\ dD} \tag{8}$$

[11] Formally, let $\tilde{F}_{a|\theta,X_0}(\cdot|\theta,x_0)$ be another distribution that differs from $F_{a|\theta,X_0}(\cdot|\theta,x_0)$ only on $(\underline{a}(\theta,x_0), a_*(\theta,x_0)) \cup (a^*(\theta,x_0), \overline{a}(\theta,x_0))$. These two distributions lead to the same probability $\Pr(\chi = 1|\theta, X = x_0, Z = z) = \Pr(a \leq a(\theta,x_0,z)|\theta, X = x_0)$ since the frontier $a(\cdot,x_0,z)$ does not depend on this conditional distribution by (5).

for $c = 1, \ldots, C-2$.

Condition (8) depends on the terms of the offered contracts as well as on the damage distribution. In contrast, it does not depend on the distribution of risk and risk aversion except through the lower bound $\underline{a}$ of risk aversion.

Interestingly, if $\underline{a}$ approaches zero, condition (8) becomes

$$\frac{t_{c+2} - t_{c+1}}{t_{c+1} - t_c} > \frac{\int_{dd_{c+2}}^{dd_{c+1}} [1 - H(D)] \, dD}{\int_{dd_{c+1}}^{dd_c} [1 - H(D)] \, dD}$$

for $c = 1, \ldots, C-2$. Applying the Mean Value Theorem gives

$$\frac{t_{c+2} - t_{c+1}}{t_{c+1} - t_c} > \kappa_{c+1} \frac{dd_{c+1} - dd_{c+2}}{dd_c - dd_{c+1}},$$

where $\kappa_{c+1} = [1 - H(D^*_{c+1})]/[1 - H(D^*_c)] > 0$ with $D^*_{c+1} \in (dd_{c+2}, dd_{c+1})$ and $D^*_c \in (dd_{c+1}, dd_c)$. In particular, because $1-H(D)$ is decreasing in D, we have $\kappa_{c+1} > 1$. Thus the increments in premia should increase proportionally more than the decrements in deductibles. This relates to a well-known property of reverse nonlinear pricing as noted by Stiglitz (1977). The next corollary formalizes this result.

Corollary: *Let A1 hold and the coverages* $(t_c, dd_c), c = 1 \ldots, C \geq 2$ *satisfy the RP condition (7). When* $\underline{a}$ *approaches zero, a necessary and sufficient condition for (8) is*

$$\frac{t_{c+2} - t_{c+1}}{|dd_{c+2} - dd_{c+1}|} > \kappa_{c+1} \frac{t_{c+1} - t_c}{|dd_{c+1} - dd_c|}, \tag{9}$$

for some $\kappa_{c+1} > 1$ *and* $c = 1, \ldots, C$.

This corollary says that the observed coverages $(t_c, dd_c), c = 1, \ldots, C$ should lie on a convex curve in the (t, dd)-space. This convexity is easily verifiable in the data. It should be noted that such a theoretical property is obtained here despite non optimal contracts and bidimensional incomplete information. See also Luo, Perrigne and Vuong (2017, 2018).

When either (8) or (9) holds, any individual whose type (θ, a) lies between the frontiers $\theta_{c,c+1}(\cdot)$ and $\theta_{c+1,c+2}(\cdot)$ chooses the coverage (t_{c+1}, dd_{c+1}), for $c = 1, \ldots, C-2$. Indeed, from Lemma 1, this individual prefers (t_{c+1}, dd_{c+1}) to (t_{c+2}, dd_{c+2}), which is preferred to (t_{c+3}, dd_{c+3}), etc. Similarly, this individual prefers (t_{c+1}, dd_{c+1}) to (t_c, dd_c), which is preferred to (t_{c-1}, dd_{c-1}), etc. Thus, a (θ, a)-individual above the $\theta_{C-1,C}(\cdot)$-frontier chooses the highest coverage (i.e., the lowest deductible) (t_C, dd_C). Figure 1 illustrates the choice among the three contracts $(t_1, dd_1) = (600, 1000)$, $(t_2, dd_2) = (850, 500)$ and $(t_3, dd_3) = (1000, 250)$, which satisfy condition (9). In contrast to the case with two coverages, insurees who are on the right of the frontier 2 versus 3 now choose (t_3, dd_3).

Our previous identification results extend to more than two contracts. Specifically, under A2, the first step that identifies the marginal distribution $F_{\theta|X}(\cdot|\cdot)$ from the observed number of accidents remains the same as before. The second step identifies the conditional distribution $F_{a|\theta,X}(\cdot|\cdot,\cdot)$ at the $C-1$ frontiers between coverages c and $c+1$ for $c = 1, \ldots, C-1$ upon introducing the choice variable χ taking values $1, \ldots, C$ and the corresponding proportions of individuals choosing coverage $c = 1, \ldots, C$. Hence, under the exclusion and support assumption A3, the distribution $F(\theta, a|X)$ is identified. As a matter of fact, A3-(ii) is stronger than necessary as it suffices that the combined variations of the $C-1$ frontiers cover the $\Theta(X) \times \mathcal{A}(X_0)$ space. Moreover, if this sufficient condition is not satisfied, $F_{a|\theta,X}(a|\theta, x)$ is identified on a larger set through the variations of the $C-1$ frontiers. Thus, having more coverages helps identify the joint distribution of types (θ, a).

In most of the empirical literature on insurance, such as in Israel (2005), Cohen and Einav (2007), and Barseghyan et al. (2013), data are collected from a single company. In this case, our results immediately apply. If data combine insurees from different firms, our approach requires that the observed coverages satisfy the revealed preference condition (7) arising from individuals' choices. This condition might not hold if switching costs are present, preventing individuals from changing to their preferred coverages. However, when insurance contracts are differentiated in other dimensions, such as vehicle replacement, uninsured motorists, or roadside assistance, the analysis can be performed by conditioning on these add-ons. It then suffices to mix the recovered conditional distributions with the proportions of individuals choosing these add-ons to obtain the joint distribution $F(\cdot, \cdot)$ of risk and risk aversion.

4 Estimation Method and Monte Carlo Study

This section presents a computationally friendly three-step nonparametric procedure for estimating the joint density $f(\theta, a|X)$ of risk θ and risk aversion a given X with support $[\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}] = [0, 1] \times [0, \overline{a}]$.[12] To simplify, we omit the covariates X_0 in $X = (X_0, Z)$, which can be entertained by conditioning our estimation procedure on X_0 through a smoothing method such as kernel estimation. Our procedure follows our identification argument as the latter is constructive. Hereafter, we consider two coverages. Let $(\chi_i, J_i, D_{1i}, \ldots, D_{J_i i}, Z_i), i = 1, \ldots, N$ be the available data where $\chi_i = c$ if individual chooses coverage $c = 1, 2$. The estimation method is implemented in a Monte Carlo study.

4.1 A Three-Step Estimation Procedure

Our estimation procedure consists in three steps:

Step 1: Estimate $f_{\theta|Z}(\cdot|\cdot)$ by constrained Generalized Method of Moments (GMM) and kernel smoothing,

Step 2: Estimate $f_{\theta|\chi,Z}(\cdot|1, \cdot)$ by adapting Step 1 and conditioning on $\chi = 1$,

Step 3: Estimate $f_{a|\theta}(\cdot|\theta)$ by plugging-in estimates of $\partial\theta(a, z)/\partial a$, $\partial\theta(a, z)/\partial z$ and $\partial \Pr(\chi = 1|\theta(a, z), z)/\partial z$.

Hereafter, we present these steps in details.

Estimation of $f_{\theta|Z}(\cdot|\cdot)$

Following the identification argument, the first step estimates the density $f_{\theta|Z}(\theta|z)$ of risk θ given Z. To fix ideas, we assume that this density does not depend on Z. The problem then reduces to estimating the mixing distribution in a Poisson mixture because the observed numbers of accidents $J_i, i = 1, \ldots, N$ are i.i.d. drawn from $\Pr(J = j) = \int_{\underline{\theta}}^{\overline{\theta}} [e^{-\theta}\theta^j/j!] f_\theta(\theta) d\theta$. The statistical literature views the estimation of the Poisson mixing density $f_\theta(\cdot)$ as a Hausdorff moment problem using the empirical (raw)

[12] The interval $[0, 1]$ is a normalization, whereas $\overline{a}$ is chosen by the analyst.

moments $\hat{\mu}_m, m = 1, \ldots, M \geq 1$

$$\hat{\mu}_m = \frac{1}{N}\sum_{i=1}^{N} J_i(J_i - 1)\ldots(J_i - m + 1) \quad \text{for } m \geq 1$$

from, e.g., Hengartner (1997). Specifically, following Talenti (1987), the estimator for $f_\theta(\theta)$ is

$$\hat{f}_\theta(\theta; \hat{\lambda}) = 1 + \hat{\lambda}_1 L_1(\theta) + \ldots + \hat{\lambda}_M L_M(\theta),$$

where $L_m(\theta) = c_{m0} + c_{m1}\theta + \ldots + c_{mm}\theta^m$ is the shifted Legendre polynomial of degree m on $[0, 1]$. The first coefficient is equal to one to satisfy $\int_0^1 \hat{f}_\theta(\theta; \lambda)d\theta = 1$, and $\hat{\lambda}_m = c_{m0} + c_{m1}\hat{\mu}_1 + \ldots + c_{mm}\hat{\mu}_m$ for $m = 1, \ldots, M$. This follows by solving the M equations

$$\int_0^1 \theta^m \hat{f}_\theta(\theta; \lambda)d\theta = \hat{\mu}_m, m = 1, \ldots, M. \tag{10}$$

Hengartner (1997) shows that the optimal convergence rate is attained when $M = \log N / \log(\log N)$. This rate is slow relative to $\sqrt{N}$, reflecting the difficulty of estimating $f_\theta(\cdot)$.[13]

To improve the finite sample properties, we impose that the resulting estimated density is nonnegative through a constrained GMM estimator.[14] Specifically, $\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_M)$ is obtained as

$$\hat{\lambda} = \text{argmin}_{\lambda=(\lambda_1,\ldots,\lambda_M)}[\hat{\mu} - \mu(\lambda)]'V^{-1}[\hat{\mu} - \mu(\lambda)],$$

subject to $\hat{f}_\theta(\theta; \lambda) \geq 0$, where $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_M)$ and $\mu(\lambda) = (\mu_1(\lambda), \ldots, \mu_M(\lambda))$ with $\mu_m(\lambda) = \int_0^1 \theta^m \hat{f}_\theta(\theta; \lambda)d\theta$. The weighting matrix is $V = \text{diag}(\widehat{\text{Var}}(\hat{\mu}_m))$, where

$$\widehat{\text{Var}}(\hat{\mu}_m) = \frac{1}{N^2}\sum_{i=1}^{N}[J_i(J_i - 1)\ldots(J_i - m + 1)]^2 - \frac{1}{N}\hat{\mu}_m^2.$$

[13] An alternative estimator consists in inverting the empirical characteristic function of θ by Fourier inversion. See Aryal, Perrigne and Vuong (2019). The deconvolution estimator implicitly requires that moments are well estimated. In the case of automobile insurance, the number of accidents tends to be small rendering estimation of moments above four or five very imprecise.

[14] We are grateful to Matheus Silva for proposing this method. See Silva (2024).

When the marginal density of θ depends on Z, we adapt the estimator by conditioning on $Z = z$ upon considering the empirical conditional moment $\hat{\mu}_m(z)$ obtained from a nonparametric regression of $J_i(J_i - 1)\dots(J_i - m + 1)$ on Z_i since $\mathrm{E}[\theta^m|Z] = \mathrm{E}[J(J - 1)\dots(J - m + 1)|Z]$. This gives $\hat{f}_{\theta|Z}(\theta|z)$ upon applying the above estimator for each z value.

Estimation of $f_{\theta|\chi,Z}(\cdot|1,\cdot)$

The third step requires an estimator of the conditional choice probability $\Pr[\chi = 1|\theta, z]$ that an individual with risk θ and covariates $Z = z$ chooses coverage 1. This probability is given by (6), whose right-hand side involves $f_{\theta|\chi,Z}(\theta|1, z)$, $\nu_1(z)$ and $f_{\theta|Z}(\theta|z)$. The term $\nu_1(z)$ is the probability of choosing coverage 1 given $Z = z$, and can be estimated by a nonparametric regression of χ_i on Z_i. The conditional density $f_{\theta|Z}(\theta|z)$ is estimated by $\hat{f}_{\theta|Z}(\theta|z)$ obtained in the first step. It remains to estimate $f_{\theta|\chi,Z}(\cdot|1, z)$. A natural method would apply the first step estimator on the subsample of individuals choosing coverage 1. However, this method suffers from a possible irregularity at $\theta(\overline{a}, z)$ when the latter belongs to $(0, 1)$ as displayed in Figure 1 or Cohen and Einav (2007, Figure 2).[15]

To address this difficulty, we note that by Bayes rule we have $f_{\theta|\chi,Z}(\theta|1, z) = f_{\theta|Z}(\theta|z)/\nu_1(z)$ when $\theta \in [0, \theta(\overline{a}, z)]$ because all such θ-individuals always choose coverage 1. Thus, from above $f_{\theta|\chi,Z}(\cdot|1, z)$ is readily estimated on $[0, \theta(\overline{a}, z)]$. It remains to estimate this density on $[\theta(\overline{a}, z), \min\{\theta(0, z), 1\}]$. We have

$$f_{\theta|\chi,Z}(\theta|1, z) = \left[1 - \frac{F_{\theta|Z}(\theta(\overline{a}, z)|z)}{\nu_1(z)}\right] g(\theta|1, z) \tag{11}$$

when $\theta \in [\theta(\overline{a}, z), \min\{\theta(0, z), 1\}]$, where $g(\cdot|1, z)$ is the conditional density of θ given $\{\theta > \theta(\overline{a}, z), \chi = 1, Z = z\}$. Thus the problem reduces to estimating the density $g(\cdot|1, z)$. The support boundaries $\theta(\overline{a}, z)$ and $\theta(0, z)$ are estimated by letting $a = \overline{a}$ and $a = 0$, respectively, in the estimated frontier

$$\hat{\theta}(a, z) = \frac{t_2(z) - t_1(z)}{\int_{dd_2(z)}^{dd_1(z)} e^{aD}[1 - \hat{H}(D|z)]dD}, \tag{12}$$

where $\hat{H}(D|Z)$ is a nonparametric estimator of the damage distribution.

[15] A formal proof is available upon request from the authors.

To estimate $g(\theta|1,z)$, we apply the first-step estimator accounting for its support $[\theta(\overline{a},z),\theta(0,z)]$.[16] Specifically, let

$$\begin{aligned} g(\theta|1,z) &= \frac{1}{\theta(0,z)-\theta(\overline{a},z)} \\ &\quad \times \left[1+\beta_{1z}L_1\left(\frac{\theta-\theta(\overline{a},z)}{\theta(0,z)-\theta(\overline{a},z)}\right)+\ldots+\beta_{Mz}L_M\left(\frac{\theta-\theta(\overline{a},z)}{\theta(0,z)-\theta(\overline{a},z)}\right)\right], \end{aligned}$$

where the coefficients $(\beta_{1z},\ldots,\beta_{Mz})$ depend on z. By (11), the mth moment of $g(\theta|1,z)$ is obtained from the mth moment of $f_{\theta|\chi,Z}(\theta|1,z)$ using

$$\mathrm{E}(\theta^m|\chi=1,z)=\int_0^{\theta(\overline{a},z)}\theta^m\frac{f_{\theta|Z}(\theta|z)}{\nu_1(z)}d\theta+\left[1-\frac{F_{\theta|Z}(\theta(\overline{a},z)|z)}{\nu_1(z)}\right]\int_{\theta(\overline{a},z)}^{\theta(0,z)}\theta^m g(\theta|1,z)d\theta,$$

for $m=1,\ldots,M$. The estimated parameters $(\hat{\beta}_{1z},\ldots,\hat{\beta}_{Mz})$ are obtained by GMM subject to the constraints

$$\begin{aligned} &(i) \quad 0\le\left[1-\frac{F_{\theta|Z}(\theta(\overline{a},z)|z)}{\nu_1(z)}\right]g(\theta|1,z)\le\frac{f_{\theta|Z}(\theta|z)}{\nu_1(z)}, \\ &(ii) \quad \left[1-\frac{F_{\theta|Z}(\theta(\overline{a},z)|z)}{\nu_1(z)}\right]g(\theta(\overline{a},z)|1,z)=\frac{f_{\theta|Z}(\theta(\overline{a},z)|z)}{\nu_1(z)}, \end{aligned}$$

upon replacing $\theta(0,z),\theta(\overline{a},z)$, $f_{\theta|Z}(\theta|z),F_{\theta|Z}(\theta|z),\nu_1(z)$ and $\mathrm{E}(\theta^m|\chi=1,z)$ by their estimated counterparts.[17] In particular, $\hat{\mathrm{E}}(\theta^m|\chi=1,z)$ is obtained from a nonparametric regression of $J_i(J_i-1)\ldots(J_i-m+1)$ on Z_i using the subsample of individuals who choose coverage 1. This gives the estimator $\hat{g}(\theta|1,z)$ and hence the estimator $\hat{f}_{\theta|\chi,z}(\theta|1,z)$ using (11). The estimator of $\Pr[\chi=1|\theta,z]$ is $\widehat{\Pr}[\chi=1|\theta,z]=\hat{f}_{\theta|\chi,z}(\theta|1,z)\hat{\nu}_1(z)/\hat{f}_{\theta|Z}(\theta|z)$.

Estimation of $f_{a|\theta}(\cdot|\cdot)$

The third step estimates the density $f_{a|\theta}(a|\theta)$ of risk aversion a, conditional on risk θ under the exclusion restriction in Assumption A3-(i). Our argument of Section 3 identifies the distribution $F_{a|\theta}(\cdot|\theta)$ for those values of a for which $a=a(\theta,z)$ for some z. This is inconvenient for estimating its density $f_{a|\theta}(\cdot|\theta)$. We exploit instead the

[16]We assume that $\theta(0,z)\le 1$ as in Figure 1 and in Cohen and Einav (2007, Figure 2).

[17]The first constraint follows from $f_{\theta|Z}(\theta|z)=f_{\theta|\chi,Z}(\theta|1,z)\nu_1(z)+f_{\theta|\chi,Z}(\theta|2,z)\nu_2(z)\ge f_{\theta|\chi,Z}(\theta|1,z)\nu_1(z)$ and (11). The second constraint imposes the continuity of $f_{\theta|\chi,Z}(\cdot|1,z)$ at $\theta(\overline{a},z)$. To improve finite sample properties, we also impose $g(\theta(0,z)|1,z)=0$ in Section 4.2.

identity $F_{a|\theta}[a(\theta, z)|\theta] = \Pr[\chi = 1|\theta, z]$. Differentiate it with respect to z and using $\partial a(\theta, z)/\partial z = -\{\partial\theta[a(\theta, z), z]/\partial z\}/\{\partial\theta[a(\theta, z), z]/\partial a\}$ give

$$f_{a|\theta}[a(\theta, z)|\theta] = -\frac{\partial\theta[a(\theta, z), z]\partial a}{\partial\theta[a(\theta, z), z]/\partial z} \times \frac{\partial \Pr[\chi = 1|\theta, z]}{\partial z}, \tag{13}$$

where $a(\cdot, z)$ is the inverse of $\theta(\cdot, z)$, and $\dim Z = 1$ to simplify. In particular, we identify $f_{a|\theta}(\cdot|\theta)$ on the range of $a(\theta, z)$ when z varies. Under the full support assumption A3-(ii), this range is $[0, \overline{a}]$. To estimate $f_{a|\theta}(\cdot|\theta)$ at a value a in the range of $a(\theta, z)$, we use numerical derivatives for $\partial\widehat{\Pr}[\chi = 1|\theta, z]/\partial z$ and $\partial\hat{\theta}(a, z)/\partial z$, while $\partial\hat{\theta}(a, z)/\partial a$ is readily available from (5), where $H(D|z)$ is replaced by its estimate $\hat{H}(D|z)$.[18] With $C > 2$ contracts, there are $C - 1$ frontiers $\theta_{c,c+1}(a, z)$, $c = 1, \ldots, C - 1$. Thus $f_{a|\theta}(\cdot|\theta)$ can be estimated on a larger range of values of a when z varies.

4.2 A Monte Carlo Study

This section implements the above estimator on simulated data.

Data-Generating Process

We consider a Monte Carlo setup that captures some basic features of automobile insurance data. Risk θ and risk aversion a are marginally distributed as Beta(2,3) on $[0, 1]$ and 10^{-3}Beta(1,3) on $[0, 10^{-3}]$, respectively. The range of values is similar to those found by Cohen and Einav (2007, Figure 1). In agreement with the intuition that risk aversion is associated with a tendency to take greater precautions, we allow for a negative association between risk and risk aversion through a Gaussian copula with correlation $\rho = -0.5$. See Finkelstein and McGarry (2006), whose empirical results support this intuition. Damages are exponentially and independently distributed with a mean of 5,000, whereas the number J of accidents is distributed as Poisson with parameter θ. We present a simplified version of Assumption A3-(i) with an exogenous variable Z that is uniformly distributed on $[100, 200]$ and independently of (θ, a). We consider two coverages with fixed deductibles at 1,000 and 500 with premia $3.25Z$ and

[18] The resulting estimator may not be positive. One can take its absolute value. When the range is the full support, normalizing it by its integral over $[0, 1]$ provides an estimator of $f_{a|\theta}(\cdot|\theta)$ satisfying the properties of a density.

700 for coverages 1 and 2, respectively. Having fixed deductibles is standard among insurance companies. Though Z is independent of (θ, a, J, D), it enters in the insuree's contract choice as he/she chooses coverage 1 if his/her risk θ_i is below the frontier $\theta(a_i, Z_i)$, where the frontier $\theta(a, Z)$ is given by (12). This frontier varies in Z through the numerator, which is sufficient for identification and estimation.

We draw a sample of 100,000 triplets (θ_i, a_i, Z_i) from $F(\theta, a)$ and $U(100, 200)$. The value of Z_i determines the pair of offered coverages and the frontier (12). Individual i chooses coverage $(1,000, 3.25Z_i)$, i.e. $\chi_i = 1$ if his/her risk $\theta_i \leq \theta(a_i, Z_i)$. Given θ_i, a number J_i of accidents is drawn from a Poisson with mean θ_i. Damages $D_{1i}, \ldots, D_{J_i i}$ are drawn from $H(\cdot)$. Figure 2 displays the observations (θ_i, a_i) for one simulated sample. The frontiers $\theta(a, z)$ when z varies from 110 (right curve) to 190 (left curve) provide the locuses of points (θ, a) for which the z-individuals are indifferent between the two coverages. If a z-individual has a (θ, a) pair under $\theta(\cdot, z)$, then he/she chooses coverage 1; otherwise, he/she chooses the second coverage offering a better protection. Figure 3 displays the histogram of the number of accidents. A large majority of individuals have no accident, and the proportions decline sharply to reach values close to zero for $J \geq 4$, in agreement with Cohen and Einav (2007, Table 2B). Using this random sample, we perform the estimation procedure detailed above. We repeat this exercise 100 times.

Monte Carlo Results

Figure 4 shows the estimated marginal density of the expected number θ of accidents. It displays the true density as well as the 90% confidence interval. The true curve is within the corresponding bounds, which are remarkably narrow. The constrained GMM estimator is implemented using $M = 4$ moments which is the integer part of $\log N/(\log \log N)$. To save space, we do not display the results of Step 2 because it is an intermediary step where we use the estimator from (11) since the frontiers $\theta(a, z)$ have an irregularity at $\theta(\overline{a}, z) \in (0, 1)$. We apply the constrained GMM estimator for the density $g(\theta|1, z)$ with $M = 4$ moments on its support $[\theta(\overline{a}, z), \theta(0, z)]$. This step also requires estimates of the probability $\nu_1(z)$ of choosing coverage 1 as well the kernel regression of $J_i(J_i - 1) \ldots (J_i - m + 1)$ on Z_i with $m = 1, \ldots, 4$. Kernel estimators are performed using rule-of-thumb bandwidths.

Figure 5 displays the density $\hat{f}_{a|\theta}(\cdot|0.4)$ conditional on $\theta = 0.4$ since θ is distributed as $B(2,3)$ with mean 0.4. This density estimator is obtained from (13). Figure 5 also provides the 90% confidence interval, which is relatively narrow and contains the true conditional density. It is worth noting that the range of $a(0.4,z)$ is $[0,10^{-3}]$ when z varies. In contrast, Figure 6 displays $\hat{f}_{a|\theta}(\cdot|0.6)$ conditional on $\theta = 0.6$. We observe that the range of $a(0.6,z)$ is $[0, 0.44\times 10^{-3}]$. This finding illustrates that the support assumption partially holds as the variation in z is not sufficient to estimate this conditional density at $\theta = 0.6$ on its full support $[0,10^{-3}]$. As discussed previously, this issue could be alleviated when observing more than two insurance options. Nonetheless, the 90% confidence interval on the identified range contains the true density except at the leftmost boundary. Boundary effects are typical in nonparametric estimation and can be corrected.

5 Conclusion

Our paper addresses the identification and estimation of insurance models where insurees have private information about their risk and risk aversion. Our model also includes random damages and the possibility of multiple accidents. Despite bunching due to multidimensional types and a finite number of offered coverages, we identify the model primitives by exploiting the observed number of claims. We also develop a nonparametric estimation procedure that is computationally friendly. Our results apply to any form of competition and do not rely on the optimality of offered coverages in contrast to the previous literature on the identification and estimation of models with private information. Thus, their optimality could be tested upon the specification of an appropriate model of market competition. Beyond optimality, several counterfactuals can be performed. For instance, we can assess the gain/loss for both parties from (i) reducing the range of insurees' characteristics that the insurer can use to discriminate insurees such as gender, age, or location, (ii) increasing the number of existing coverages and/or changing their terms, and (iii) implementing other coverages than premium/deductible with (say) a proportional deductible.

In terms of future lines of research, first, our results extend to a broad range of insurance data, such as in health, provided the analyst observes repeated outcomes, e.g. insurees' claims. In particular, we may want to extend our identification results to allow for some form of moral hazard. Second, in the case of automobile insurance, we could endogenize the car choice given insuree's risk and risk aversion. This extension would lead to a model explaining the car choice, the coverage choice, the number of accidents, and the damages. Third, several existing data sets on automobile and home insurance used by Israel (2005), Cohen and Einav (2007), Sydnor (2010), and Barseghyan, Molinari, O'Donoghue and Teitelbaum (2013) could be analyzed using our empirical framework.

Appendix

The appendix contains the extension to health insurance as well as the proofs of Lemmas 1 and 2.

The Case of Health Insurance: Up to some variations, health insurance involves a premium t as well as a per period deductible dd and a copayment γ per (say) medical visit. In particular, the deductible is not per visit, while the copayment arises on the first visit after the deductible is met. In this case, when buying a contract (t, dd, γ), the (θ, a)-patient has an expected utility

$$V(t, dd, \gamma; \theta, a, w) = -e^{-a(w-t)}\mathrm{E}[e^{-aY(dd,\gamma)}|\theta],$$

where the insuree's expense beyond the premium (commonly referred as out-of-pocket) is $Y(dd, \gamma) = (D_1 + \ldots + D_J)1\!\!1(D_1 + \ldots + D_J \leq dd) + (dd + (J - J^\dagger)\gamma)1\!\!1(D_1 + \ldots + D_J > dd)$ with J the number of visits and $J^\dagger$ the number of visits at which the deductible is met, i.e., $J^\dagger = \mathrm{argmin}_{j=1,\ldots,J} D_1 + \ldots + D_j > dd$. The expectation is with respect to the total expense $D_1 + \ldots + D_J$ and the number J of visits which depends on θ. In the case of health coverage, the per visit expenses $D_j, j = 1, \ldots, J$ may be viewed as independent conditional on the patient's health conditions such as cancer, diabetes, which are observed by the analyst. Similarly, conditioning on the patient's health conditions alleviates the possible dependence between D_j and the expected number of medical visits θ. See Assumption A2 in the text for the introduction of insurees' characteristics. Letting $m(dd, \gamma; \theta, a) = \mathrm{E}[e^{-aY(dd,\gamma)}|\theta]$, the certainty equivalent becomes

$$CE(t, dd, \gamma; \theta, a, w) = w - t - \frac{\log m(dd, \gamma; \theta, a)}{a},$$

which is similar to (2).

Proof of Lemma 1: The first part of (i) is immediate from (2) since $\phi_a(dd) = 1$ when $dd = 0$. When $dd > 0$, the derivative of (2) with respect to θ is $-(\phi_a - 1)/a$. Since $\phi_a > 1$, $CE(t, dd; \theta, a, w)$ is decreasing in θ. For the derivative of (2) with respect to a, we note that

$$\phi_a(dd) = \int_0^{dd} e^{aD} dH(D) + e^{add}[1 - H(dd)] = 1 + a\int_0^{dd} e^{aD}[1 - H(D)]dD \qquad \text{(A.1)}$$

by integration by parts. Thus (2) gives

$$CE(t, dd; \theta, a, w) = w - t - \theta\int_0^{dd} e^{aD}[1 - H(D)]dD.$$

Hence, $CE(t, dd; \theta, a, w)$ is decreasing in a.

We now prove (ii). The frontier between $\mathcal{C}_1$ and $\mathcal{C}_2$ is defined as the locus of (θ, a)-insurees who are indifferent between the two coverages, i.e., for whom $CE(t_1, dd_1; \theta, a, w) = CE(t_2, dd_2; \theta, a, w)$. Using (2) this gives

$$t_1 + \frac{\theta[\phi_a(dd_1) - 1]}{a} = t_2 + \frac{\theta[\phi_a(dd_2) - 1]}{a}.$$

Solving for θ as a function of a gives (3) upon using (A.1). Moreover, from (3) it is easy to see that $\theta(a)$ decreases in a. The last part of (ii) also follows as $CE(t_1, dd_1; \theta, a, w) > CE(t_2, dd_2; \theta, a, w)$ if and only if $\theta < \theta(a)$. □

Proof of Lemma 2: Fix $c = 0, 1, \ldots, C-1$. From (3), the frontier $\theta_{c,c+1}(\cdot)$ between coverages (t_c, dd_c) and $(t_{c+1},\, dd_{c+1})$ is given by

$$\theta_{c,c+1}(a) \quad = \quad \frac{a(t_{c+1} - t_c)}{\phi_a(dd_c) - \phi_a(dd_{c+1})} = \frac{t_{c+1} - t_c}{\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD}.$$

Thus, for any $c = 0, 1, \ldots, C-2$, $\theta_{c,c+1}(\cdot) < \theta_{c+1,c+2}(\cdot)$ on $a \in [0, \overline{a}]$ if and only if

$$\frac{t_{c+2} - t_{c+1}}{t_{c+1} - t_c} > \frac{\int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD}{\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD} \tag{A.2}$$

for all $a \in [\underline{a}, \overline{a}]$. For any such c, we first show that the RHS of (A.2) decreases in $a \in (0, +\infty)$. Adding 1 to the inverse of the RHS, it is equivalent to showing that the ratio $\int_{dd_{c+2}}^{dd_c} e^{aD}[1 - H(D)]dD / \int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD$ is increasing in a, i.e., that $\int_{dd_{c+2}}^{dd} e^{aD}[1 - H(D)]dD$ is log-supermodular in $(a, dd) \in (0, +\infty) \times (dd_{c+2}, \overline{d})$ given dd_{c+2} since $d_{c+1} < d_c$.[19] The latter holds by Lemma C.1 upon letting $x = a$, $y = dd$, $\overline{y} = \overline{d}$ and $y_\dagger = dd_{c+2}$.

We now show that condition (8) is necessary and sufficient for $\theta_{c,c+1}(\cdot) < \theta_{c+1,c+2}(\cdot)$ on $[\underline{a}, \overline{a}]$. Because (A.2) must hold at $\underline{a}$, then (8) is necessary. Since the RHS of (A.2) decreases in $a \in (0, +\infty)$, it is bounded above for all $a \in [\underline{a}, \overline{a}]$ by the RHS of (A.2) evaluated at $\underline{a}$. Thus (8) implies that (A.2) holds for all $a \in [\underline{a}, \overline{a}]$ thereby establishing sufficiency. □

[19] A positive function $f(x, y)$ is log-supermodular if $\log f(x, y)$ is supermodular.

References

Armstrong, Mark (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64, 51-75.

Aryal, Gaurab and **Federico Zincenko** (2023): "Identification and Estimation of Multidimensional Screening," Unpublished manuscript, University of Washington St Louis.

Aryal, Gaurab, **Isabelle Perrigne**, and **Quang Vuong** (2016): "Identification of Insurance Models with Multidimensional Screening," Unpublished manuscript, New York University.

Aryal, Gaurab, **Isabelle Perrigne**, and **Quang Vuong** (2019): "Econometrics of Insurance Models with Multidimensional Types," Unpublished manuscript, New York University.

Aryal, Gaurab, **Isabelle Perrigne**, **Quang Vuong** and **Haiqing Xu** (2024): "Supplement to 'Econometrics of Insurance with Multidimensional Types'," *Quantitative Economics Supplemental Material*.

Barseghyan, Levon, **Francesca Molinari**, **Ted O'Donoghue** and **Joshua Teitelbaum** (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103, 2499-2529.

Berry, Steven and **Philip Haile** (2014): "Identification in Differentiated Product Markets," *Econometrica*, 82, 1749-1797.

Billingsley, Patrick (1995): *Probability and Measure*, 3rd Edition, New York, USA: Academic Press.

Chen, Xiaohong,**Matthew Gentry**, **Tong Li** and **Jingfeng Lu** (2023): "Identification and Inference in First-Price Auctions with Risk-Averse Bidders and Selective Entry," Unpublished manuscript, Yale University.

Chiappori, Pierre-André and **Bernard Salanié** (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56-78.

Cohen, Alma and **Liran Einav** (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745-788.

Cohen, Alma and **Peter Siegelman** (2010): "Testing for Adverse Selection in Insurance Markets," *Journal of Risk and Insurance*, 77, 39-84.

Crawford, Gregory and **Matthew Shum** (2007): "Monopoly Quality Choice in Cable Television," *Journal of Law and Economics*, 50, 181-209.

Cutler, David, **Amy Finkelstein**, **Kathleen McGarry** (2008): "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," *American Economic Review, Papers and Proceedings*, 98, 157-162.

Einav, Liran and **Amy Finkelstein** (2011): "Selection in Insurance Markets: Theory and Empirics in Pictures," *Journal of Economic Perspectives*, 25, 115-138.

Fang, Hanming, **Michael Keane** and **Daniel Silverman** (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 116, 303-350.

Finkelstein, Amy and **Kathleen McGarry** (2006): "Multiple Dimensions of Private Information: Evidence from the Long-term Care Insurance Market," *American Economic Review*, 96, 938-958.

Gayle, George-Levi and **Robert Miller** (2015): "Identifying and Testing Models of Managerial Compensation," *Review of Economic Studies*, 82, 1074-1118.

Gentry, Matthew and **Tong Li** (2014): "Identification in Auctions with Selective Entry," *Econometrica*, 82, 315-344.

Guerre, Emmanuel, **Isabelle Perrigne** and **Quang Vuong** (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525-574.

Greenwood, Major and **G. Udny Yule** (1920): "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journal of the Royal Statistical Society, Series A*, 83, 255-279.

Hengartner, Nicolas W. (2017): "Adaptive Demixing in Poisson Mixture Models," *Annals of Statistics*, 25, 917-928.

Imbens, Guido and **Whitney Newey** (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481-1512.

Israel, Mark (2005): "Services as Experience Goods: An Empirical Examination of Consumer Learning in Automobile Insurance," *American Economic Review*, 95, 1444-1463.

Kong, Yunmi (2018): "Selectice Entry in Auctions: Estimation and Evidence," Unpublished manuscript, Rice University.

Kong, Yunmi, **Isabelle Perrigne** and **Quang Vuong** (2022): "Multidimensional Auctions of Contracts: An Empirical Analysis," *American Economic Review*, 112, 1703-1736.

Luo, Yao, **Isabelle Perrigne** and **Quang Vuong** (2017): "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS," Unpublished manuscript, New York University.

Luo, Yao, **Isabelle Perrigne** and **Quang Vuong** (2018): "Structural Nonlinear Pricing," *Journal of Political Economy*, 126, 2523-2568.

Marmer, Vadim and **Artyom Shneyerov** (2012): "Quantile Based Nonparametric Inference for First-Price Auctions," *Journal of Econometrics*, 167, 345-357.

Matzkin, Rosa (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

Rao, B.L.S. Prakasa (1992): *Identifiability in Stochastic Models: Characterization of Probability Distributions*, New York, USA: Academic Press.

Rochet, Jean-Charles and **Philippe Chone** (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 66, 783-826.

Rochet, Jean-Charles and **Lars Stole** (2003): "The Economics of Multidimensional Screening," in *Advances in Eonomic Theory, 8th World Congress*, ed. by M. Dewatripont, L.P. Hansen and S. Turnovsky, Cambridge, UK: Cambridge University Press, 150-197.

Rothschild, Michael and **Joseph Stiglitz** (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 236-257.

Shohat, James and **Jacob Tamarkin** (1943): *The Problem of Moments*, Providence, USA: American Mathematical Society.

Silva, Matheus (2024): *Essays in Econometrics*, PhD dissertation, New York University.

Stiglitz, Joseph (1977): "Monopoly, Nonlinear Pricing and Incomplete Information: The Insurance Market," *Review of Economic Studies*, 44, 407-430.

Sydnor, Justin (2010): "(Over)Insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177-199.

Talenti, Giorgio (1987): "Recovering a Function from a Finite Number of Moments," *Inverse Problems*, 3.

Figure 1: *$C = 2$ and $C = 3$ Coverages*

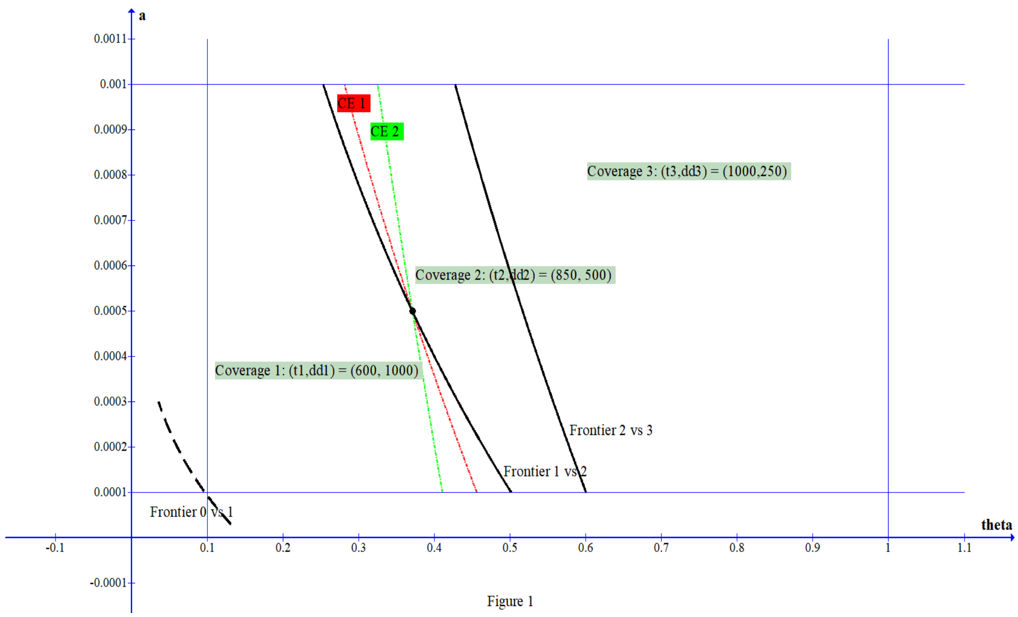

Figure 2: *Scatter Plot of (a_i, θ_i)*

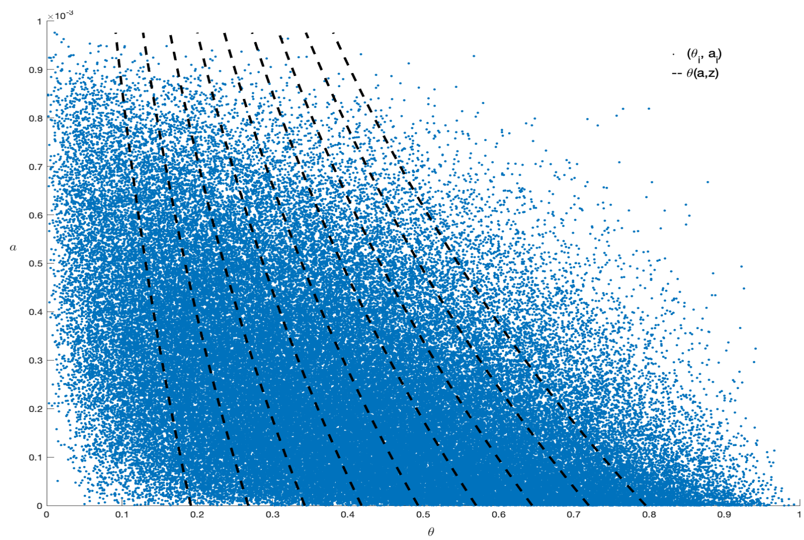

Figure 3: *Distribution of the Number of Accidents*

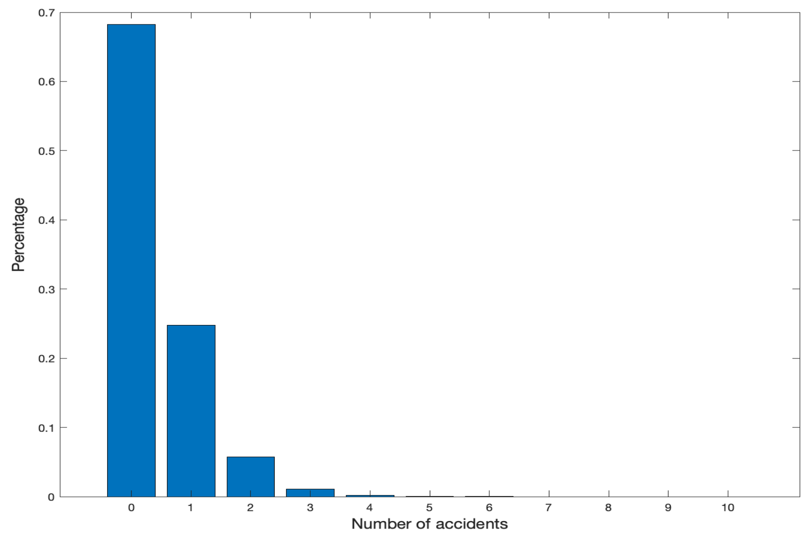

Figure 4: *Estimated Risk Density* $\hat{f}_{\theta}(\cdot)$

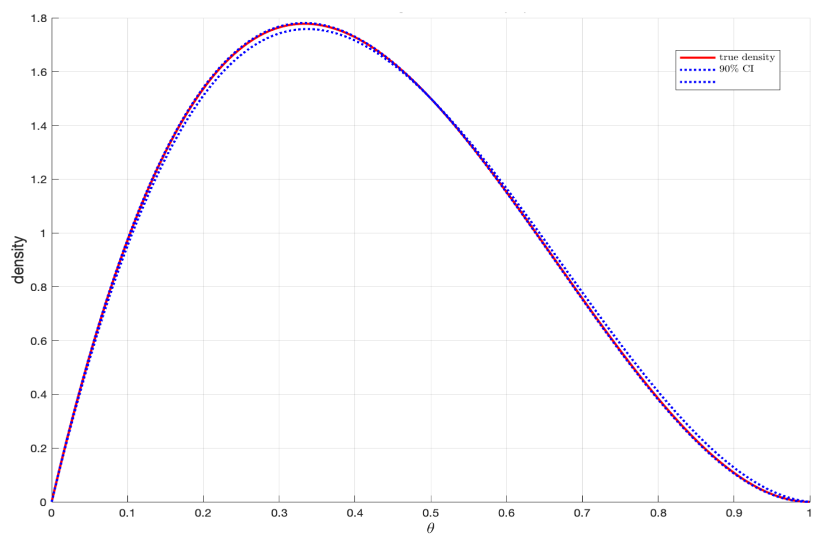

Figure 5: $\hat{f}_{a|\theta}(\cdot|0.4)$

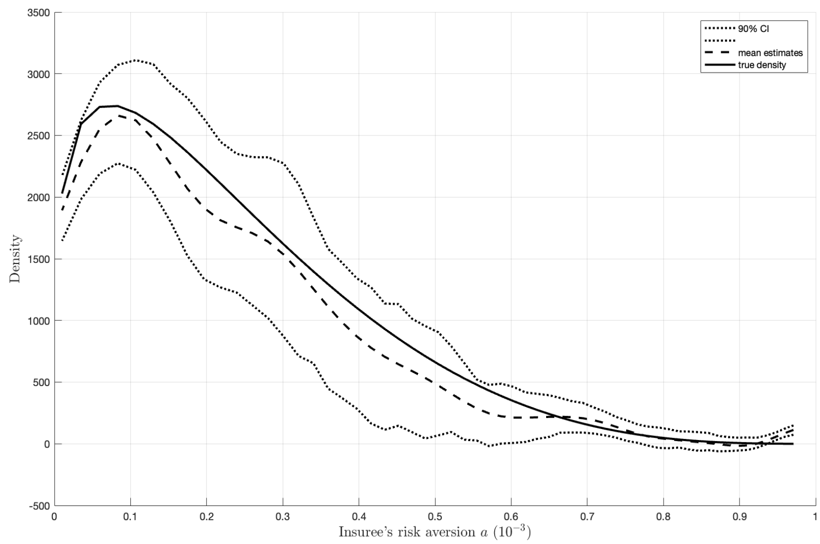

Figure 6: $\hat{f}_{a|\theta}(\cdot|0.6)$