Identification of Insurance Models with Multidimensional Screening

Gaurab Aryal
Pennsylvania State University

Isabelle Perrigne
Pennsylvania State University

Quang Vuong
Pennsylvania State University

September 2009

Correspondence to: I. Perrigne, Department of Economics, Pennsylvania State University, University Park, PA 16802. e-mail address : perrigne@psu.edu.

Abstract

We study the identification of an insurance model with multidimensional screening, where insurees are characterized by risk and risk aversion. The model is solved using the concept of certainty equivalence under constant absolute risk aversion and an unspecified joint distriubtion of risk and risk aversion. The paper then analyzes how data availability constraints identification under four data scenarios from the ideal situation to a more realistic one. The observed number of accidents for each insuree plays a key role to identify the model.

In a first part, we consider the case of a continuum of coverages offered to each insuree whether the damage distribution is fully observed or truncated. Truncation arises from that an insuree files a claim only when the accident involves a damage above the deductible. Despite bunching due to multidimensional screening, we show that the joint distribution of risk and risk aversion is identified. In a second part, we consider the case of a finite number of coverages offered to each insuree. When the full damage distribution is observed, we show that despite additional pooling due to the finite number of contracts, the joint distribution of risk and risk aversion is identified under a full support assumption and a conditional independence assumption involving the car characteristics. When the damage distribution is truncated, the joint distribution is identified up to the probability that the damage is above the deductible. In a third part, we derive the restrictions imposed by the model on observables for the fourth scenario. We also propose several identification strategies for the damage probability at the deductible. These identification results are further exploited in a companion paper developing an estimation method with an application to insurance data.

Keywords: Nonparametric Identification, Multidimensional Screening, Insurance, Moment Generating Function.

Identification of Insurance Models with Multidimensional Screening

G. Aryal, I. Perrigne & Q. Vuong

1 Introduction

Identification of structural models in industrial organization has received much attention over the past fifteeen years. See the survey by Athey and Haile (2007) on the nonparametric identification of auction models. The problem of identification in econometrics has a long history. See Koopmans (1949) and Hurwicz (1950). It is a key step for the econometric and empirical analysis of structural models. For instance, the labor literature provides many examples of the role played by identification in empirical studies as discussed by Heckman (2001). Studies on identification have known a renewed interest due to the development of nonparametric models with nonseparable error terms (see Matzkin (1994, 2007)), and to the use of structural models in empirical industrial organization. The problem of (nonparametric) identification is important for several reasons. First, it allows to assess the conditions required (if any) to recover uniquely the structure of the model from the observables while minimizing parametric assumptions. Second, it highlights which variations in the data allows one to identify each element of the structure. Third, some important questions related to the structural analysis of models can be addressed once identification is established. One can think of which distribution of the data can be rationalized by the model, or what restrictions the model imposes on the observables that can be used to test the model validity.

More recently, the identification of several models with incomplete information has been addressed. Several lessons can be drawn. First, the optimal behavior of economic agents plays an important role in identifying the model. For instance, in nonlinear pricing models, the optimality of the tariff offered to consumers needs to be considered in addition to the optimal consumers' behavior to recover the latter's willingness-to-pay distribution and marginal utilities. See Perrigne and Vuong (2009). In this case, the first-order conditions play a crucial role in establishing identification. Second, identification can be achieved with instrumental variables and exclusion restrictions, which have been widely used in the early literature on identification. Third, the one-to-one mapping between the unobserved agent's private information and the observed outcome such as the bidder's private value and his bid in auctions is a key element on which identification relies. See Guerre, Perrigne and Vuong (2000) and Athey and Haile (2007) in the context of auctions.

Our paper differs from the previous literature in several dimensions. First, we consider a model with multidimensional screening in which bunching/pooling cannot be avoided. In this case, identification cannot rely exclusively on the one-to-one mapping between the agent's unobserved types and his observed outcome/action.[1] Second, we consider a finite number of options/contracts offered to each agent, while agents' types are distributed over a continuum. In addition to the bunching arising from multidimensional screening, additional bunching arises because a finite number of contracts is offered to each agent. This represents an additional challenge in the study of identification.[2]

In this paper, we are interested in the identification of insurance models with multidimensional screening. Recent empirical studies on insurance by Cohen and Einav (2007) and Einav, Finkelstein and Schrimpf (2009) have shown an important heterogeneity in risk preferences, which may counterbalance the traditional intuition behind the Roth-

[1]Relying on Rochet and Chone (1998), Pioner (2007) addresses the identification of multidimensional screening models in a nonlinear pricing context but assumes that one of the two agent's types is observed by the analyst.

[2]Crawford and Shum (2007) consider two contracts while agents' types can take only two values thereby avoiding any bunching. Gayle and Miller (2008) adopt a similar strategy. Leslie (2004) entertains a finite number of price options through a discrete choice model to analyze consumers' behavior but takes the price schedule as exogenous. On the other hand, Perrigne and Vuong (2008, 2009) and D'Haultfoeuille and Fevrier (2007) consider a continuum of contracts in principal-agent settings.

schild and Stiglitz (1977) model of insurance. Namely, a low risk driver may buy a high coverage because of high risk aversion and conversely. Thus, a model of insurance needs to incorportate an additional component of asymmetric information, i.e. agent's risk aversion. Multidimensional screening, however, is known to be a difficult theoretical problem because of the violation of the Spence-Mirrlees (single-crossing) condition. Thus, bunching will arise. See Rochet and Stole (2003) for a survey. In our case, following Aryal and Perrigne (2009), this problem is solved using the certainty equivalence for no coverage. In particular, the latter allows one to separate insurees though each level of certainty equivalence corresponds to a set of individuals with different risk and risk aversion. The model structure is given by the damage distribution and the joint distribution of risk and risk aversion. For convenience, we consider constant absolute risk aversion as the latter leads to an explicit expression for the certainty equivalence.

We proceed as follows. We consider several data scenarios from the ideal case with a continuum of contracts offered to each insuree and a fully observed damage distribution to the more realistic case with a finite number of contracts offered to each insuree and a truncated damage distribution as an insuree files a claim only if the damage is above the deductible. This allows us to better understand the role played by the data and in particular how data constraints or limits identification of primitives. Moreover, this allows us to assess which identifying assumption is needed.

The first data scenario is in the spirit of the auction literature as we exploit the one-to-one mapping between the level of certainty equivalence and the deductible to identify the distribution of certainty equivalence. The repetition of some outcome by the agent, namely the number of accidents, then plays a crucial role in identifying the joint distribution of risk and risk aversion. This contrasts with Chiappori and Salanie (2000) test of asymmetric information in automobile insurance, which relies on whether the insuree has an accident. When considering heterogeneity in risk aversion, our results show that we need to exploit the number of accidents to achieve identification. The second data scenario maintains a continuum of contracts but considers a damage distribution truncated at the deductible. Because a continuum of contracts is offered, the subpopulation choosing full insurance, i.e. a zero deductible, identifies the damage distribution and the argument of the first case applies.

When considering a finite number of contracts, identification becomes more complex as the FOCs no longer provide a one-to-one mapping between the contract terms and the insuree's private information. Though the context is different, the number of accidents plays a key role again in identifying the marginal distribution of risk. Regarding the identification of the joint distribution of risk and risk aversion, we exploit an exclusion restriction and a full support assumption requiring sufficient variations in the car characteristics. Under these assumptions, the structure is identified when the damage distribution is fully observed. On the other hand, when the damage distribution is truncated at the deductible, we obtain identification of the structure up to the knowledge of the probability that the damage is below the deductible. The latter probability is not identified. To complete these results, we derive the model restrictions on the observables in the fourth data scenario. We also explore some identifying assumptions for the probability of damage below the deductible. We consider a parameterization of the damage distribution, additional data and a set identification strategy leading to some bounds for the model structure.

The outline of the paper is as follows. Section 2 presents the model with a continuum of contracts offered to each insuree and an extension to two contracts offered. Section 3 addresses identification when a continuum of contracts is offered whether the damage distribution is fully observed or truncated at the deductible. Section 4 studies identification when only two contracts are offered to each insuree making again the distinction between a fully observed damage distribution and a truncated one. Section 5 derives the restrictions imposed by the model under the latter data scenario, while Section 6 discusses some identifying strategies for the damage probability below the deductible. Section 7 concludes.

2 A Model of Insurance

This section relies on the theoretical results of Aryal and Perrigne (2009), who solve the bidimensional screening problem in insurance. The basic idea is to use the concept of certainty equivalence to rank insurees and reduce the bidimensional screening problem into a single dimension. As expected, there is some pooling at equilibrium as agents with the same level of certainty equivalence when no insurance is bought choose the same pair of premium and deductible. They show that using certainty equivalence is not suboptimal for the insurer to screen insurees. They also derive the first-order conditions that must satisfy the premium and deductible when a continuum of coverages is offered and when a finite number of coverages is offered. In this section, we briefly review the notations and results that are needed to study the identification of the model. A notable difference of our model with theirs is the definition of risk. In the theoretical literature on insurance starting with Rotschild and Stiglitz (1976) and Stiglitz (1977), the insuree's risk is defined as the probability of accident. With such a definition of risk, Aryal, Perrigne and Vuong (2009) show that the model is not identified even in the best data scenario of a continuum of contracts and a fully observed damage distribution. Intuitively, one can identify the distribution of certainty equivalence but the nonavailability of the number of accidents for each insuree leads to the nonidentification of the joint distribution of risk and risk aversion. Because we exploit here the observed number of accidents for each insuree, for convenience we measure the insuree's risk as the expected number of accidents. From an empirical perspective, this measure makes sense as the insurer cares about the number of accidents for each insuree as each accident may involve some payment. The theoretical results of Aryal and Perrigne (2009) extend to this case.

We first introduce some notations and assumptions. Each insuree is characterized by a pair (θ, a), where θ is his risk measured as the expected number of accidents and a is his coefficient of constant absolute risk aversion (CARA). This information is known only to the insuree leading to a problem of imperfect information for the insurer. The latter is assumed to be a monopolist as in Stiglitz (1977). In contrast, in the pioneering Rothschild and Stiglitz (1976) model, insurees vary in risk only, while their risk aversion is common and known to the insurer. The pair (θ, a) is distributed as $F(\cdot,\cdot)$, which is twice continuously differentiable on its support $\Theta \times \mathcal{A} = [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}]$. The pairs (θ, a) are assumed to be independent across insurees. The insuree's utility function is assumed to be CARA, i.e. $U_a(x) = -e^{-ax}$. Each insuree may be involved in J accidents over the contract period, where J follows a Poisson distribution with parameter θ. The number of accidents J is independent across individuals. Each accident involves a damage $D_j, j = 1, \ldots, J$, which are i.i.d as $H(\cdot)$ with support $[0, \overline{d}] \subset I\!R_+$. Damages are independent of (θ, a).

Certainty Equivalence

We introduce the concept of certainty equivalence when the individual has no coverage and when he buys an insurance contract (t, dd), where t is the premium and dd the deductible. Denoting w the insuree's wealth and $p_j = \Pr[j \text{ accidents occur}] = e^{-\theta}\theta^j/j!$, the expected utility of a (θ, a) insuree without insurance is

$$\begin{aligned} V(0,0;\theta,a) &= p_0 U_a(w) + p_1 \mathrm{E}[U_a(w - D_1)] + p_2 \mathrm{E}[U_a(w - D_1 - D_2)] + \ldots \\ &= -p_0 e^{-aw} - p_1 e^{-aw}\mathrm{E}[e^{aD_1}] - p_2 e^{-aw}\mathrm{E}[e^{aD_1}]\mathrm{E}[e^{aD_2}] - \ldots \\ &= -e^{-aw}\left[p_0 + p_1\phi_a + p_2\phi_a^2 + \ldots\right] \\ &= -e^{-aw}e^{-\theta}\left(1 + \frac{\theta\phi_a}{1!} + \frac{\theta^2\phi_a^2}{2!} + \ldots\right) \\ &= -e^{-aw+\theta(\phi_a - 1)}, \end{aligned}$$

where $\phi_a = \mathrm{E}[e^{aD_1}] > 1$. The certainty equivalence $CE(0,0;\theta,a)$ of no insurance coverage is defined by the amount of certain wealth for the insuree that will give him the same level of utility when he has no coverage, i.e. by $-e^{-aCE(0,0;\theta,a)} = -e^{-aw+\theta(\phi_a-1)}$. Thus,

$$s \equiv CE(0,0;\theta,a) = w - \frac{\theta(\phi_a - 1)}{a}. \tag{1}$$

We can verify that $\partial s/\partial\theta < 0$ and $\partial s/\partial a < 0$. The certainty equivalence of no insurance coverage decreases in both risk and risk aversion. As s is a function of (θ, a), it is random and distributed as $K(\cdot)$ on $[\underline{s}, \overline{s}]$, where $\underline{s}$ corresponds to the insuree $(\overline{\theta}, \overline{a})$ and $\overline{s}$ to the insuree $(\underline{\theta}, \underline{a})$, respectively. The certainty equivalence of no insurance coverage defines a locus of pairs (θ, a) on a downward sloping curve $\theta(a)$ at s given.

We now turn to the certainty equivalence when the individual buys the insurance coverage (t, dd). The (θ, a) insuree's expected utility needs to incorporate that the damage is covered by the insurer when it is above the deductible. Thus, his utility is affected by the damage only when it is below the deductible. Using the same derivation as above where w and D_j are replaced by $w - t$ and $\min(dd, D_j)$, respectively, we obtain

$$V(t, dd; \theta, a) = -\exp\left[-a(w-t) + \theta\left(\phi_a^* - 1\right)\right],$$

where $\phi_a^* = \mathrm{E}[e^{a\min(dd,D)}] = \int e^{a\min(dd,D)} dH(D) = \int_0^{dd} e^{aD} dH(D) + e^{add}(1 - H(dd))$. The certainty equivalence for purchasing the coverage (t, dd) is given by

$$CE(t, dd; \theta, a) = w - t - \frac{\theta\left(\int_0^{dd} e^{aD} dH(D) + e^{add}(1 - H(dd)) - 1\right)}{a}. \qquad (2)$$

THE INSURER'S PROFIT

We first assume that the insurer offers a continuum of contracts $(t(\theta, a), dd(\theta, a))$, $(\theta, a) \in \Theta \times \mathcal{A}$. Under incomplete information, the insurer's expected profit is given by

$$\begin{aligned}
\mathrm{E}[\pi] &= \int_{\Theta\times\mathcal{A}} \Bigg\{ t(\theta, a) - p_1(\theta)\left[\int_0^{\overline{d}} \max(0, D_1 - dd(\theta, a)) dH(D_1)\right] \\
&\quad - p_2(\theta)\left[\int_0^{\overline{d}} \max(0, D_1 - dd(\theta, a)) dH(D_1) + \int_0^{\overline{d}} \max(0, D_2 - dd(\theta, a)) dH(D_2)\right] \\
&\quad - \dots \Bigg\} dF(\theta, a) \\
&= \int_{\Theta\times\mathcal{A}} \left[t(\theta, a) - \sum_{j=1}^{\infty} p_j(\theta) j \int_0^{\overline{d}} \max(0, D - dd(\theta, a)) dH(D)\right] dF(\theta, a) \\
&= \int_{\Theta\times\mathcal{A}} \left[t(\theta, a) - \theta \int_{dd(\theta,a)}^{\overline{d}} (1 - H(D)) dD \right] dF(\theta, a), \qquad (3)
\end{aligned}$$

where $\max(0, d - dd(\theta, a))$ reflects that the insurer covers the damage above the deductible only and $(t(\theta, a), dd(\theta, a))$ indicates the dependence of the premium and deductible on the insuree's type. The notation $p_j(\theta)$ emphasizes its dependence on the insuree's risk θ. The last equality follows from $\sum_{j=1}^{\infty} p_j(\theta) j = \theta$ and $\int_0^{\overline{d}} \max\{0, D - dd(\theta, a)\} dH(D) = \int_{dd(\theta,a)}^{\overline{d}} (1 - H(D)) dD$.

Following Aryal and Perrigne (2009), we can equivalently express the insurer's expected profit in terms of the certainty equivalence of no insurance $s = CE(0,0;\theta,a)$. In particular, the insurer does as well by proposing the same contract $(t(s), dd(s))$ for all insurees with (θ, a) pairs leading to the certainty equivalence s. Thus, $t(\theta,a) = t(s)$ and $dd(\theta,a) = dd(s)$. By making the change of variable (θ,a) to (θ,s) in (3) gives

$$\mathrm{E}[\pi] = \int_{\underline{s}}^{\overline{s}} \left[t(s) - \mathrm{E}(\theta|s) \int_{dd(s)}^{\overline{d}} (1 - H(D))dD \right] k(s)ds,$$

where $k(\cdot)$ is the density of certainty equivalence.

THE OPTIMIZATION PROBLEM

Hereafter, we solve the problem in terms of s. The contracts need to guarantee the insuree's participation and his true type revelation. For the latter, we have

$$\max_{\tilde{s}\in[\underline{s},\overline{s}]} CE(t(\tilde{s}), dd(\tilde{s}); \theta, a) = \max_{\tilde{s}\in[\underline{s},\overline{s}]} w - t(\tilde{s}) - \frac{\theta \left[\int_0^{dd(\tilde{s})} e^{aD} dH(D) + e^{add(\tilde{s})}(1 - H(dd(\tilde{s}))) - 1 \right]}{a},$$

leading to the first-order condition at $\tilde{s} = s$

$$dd'(s) = -\eta(s, a, dd(s))t'(s),$$

for all $s \in [\underline{s}, \overline{s}]$, where

$$\eta(s, a, dd(s)) = \frac{\phi_a - 1}{a(w - s)e^{add(s)}[1 - H(dd(s))]}, \tag{4}$$

since $\theta = a(w - s)/(\phi_a - 1)$. This provides the incentive compatibility constraint for the insurer's optimization problem. Regarding the individual rationality constraint, Aryal and Perrigne (2009) show that (i) there is no countervailing incentives problem and (ii) it reduces to the boundary condition that sets the certainty equivalence for purchasing coverage $CE(t(\overline{s}), dd(\overline{s}); \underline{\theta}, \underline{a})$ for the $(\underline{\theta}, \underline{a})$ insuree at $\overline{s} \equiv CE(0, 0; \underline{\theta}, \underline{a})$.

Aryal and Perrigne (2009) show that the insurer's problem can be solved along the path $a(s)$, which is determined as the intersection of the insuree's (IC) constraint, i.e. $\theta = -t'(s)\exp(-add(s))/[dd'(s)(1 - H(dd(s)))]$ and his certainty equivalence s, i.e. $\theta = a(w - s)/(\phi_a - 1)$. Thus, (4) can be written as $\eta(s, a(s), dd(s))$. The Hamiltonian of the insurer's optimization problem can be written as

$$\begin{aligned} H(t(s), dd(s)) &= \left[t(s) - \mathrm{E}(\theta|s) \int_{dd(s)}^{\overline{d}} (1 - H(D)) dD\right] k(s) \\ &\quad + v(s)t'(s) + w(s)dd'(s) + r(s)\left[dd'(s) + \eta(s, a(s), dd(s))t'(s)\right], \end{aligned}$$

where $t(s)$ and $dd(s)$ are the state variables, $t'(s)$ and $dd'(s)$ are the control variables, $v(s)$, $w(s)$ and $r(s)$ are the co-state variables. Solving for the first-order conditions, $(t(s), dd(s))$ is solution of

$$\begin{aligned} &\eta(s, a(s), dd(s))\mathrm{E}(\theta|s)[1 - H(dd(s))] \\ &\quad + \frac{K(s)}{k(s)} \frac{1}{\eta(s, a(s), dd(s))} \left[-\frac{\partial \eta(s, a(s), dd(s))}{\partial dd} dd'(s) + \eta'(s, a(s), dd(s))\right] = 1, \quad (5) \\ &dd'(s) = -\eta(s, a^+(s), dd)t'(s), \quad (6) \end{aligned}$$

where $\eta'(s, a(s), dd(s))$ denotes the total derivative of $\eta(s, a(s), dd(s))$ with respect to s, with the initial condition $CE(t(dd(\overline{s})), dd(\overline{s}); \overline{s}) = \overline{s}$. See Aryal and Perrigne (2009) for the derivation of (5) and (6) interpreting their θ as the expected number of accidents. At equilibrium, a lower value of s implies more insurance, i.e. a lower deductible and a higher premium. At $\underline{s}$, we have full insurance with $dd(\underline{s}) = 0$.

Finite Number of Contracts

In practice, the principal offers a finite number C of contracts from which the agent can choose. In insurance, we observe in general two to five pairs of premium and deductible offered. To simplify the presentation, we consider $C = 2$. Our model takes C as exogenous. Let (t_1, dd_1) and (t_2, dd_2) with $t_1 < t_2$ and $dd_1 > dd_2$ be the two contracts offered by the insurer. We show how the insurer can determine these two contracts optimally. Intuitively, in addition to the pooling of pairs (θ, a) leading to the same certainty equivalence s, there will be bunching of agents with different values of s. The idea is then to determine two subsets $\mathcal{A}_1$ and $\mathcal{A}_2$ that partition $\Theta \times \mathcal{A}$ such that individuals in $\mathcal{A}_1$ and $\mathcal{A}_2$ choose (t_1, dd_1) and (t_2, dd_2), respectively.

The frontier between $\mathcal{A}_1$ and $\mathcal{A}_2$ is determined by the locus of (θ, a) insurees who are indifferent between the two contracts, i.e. for which $CE(t_1, dd_1; \theta, a) = CE(t_2, dd_2; \theta, a)$.

Using the previous expressions for certainty equivalence, the frontier is the part lying in $\Theta \times \mathcal{A}$ of the strictly decreasing curve defined by

$$
\begin{aligned}
\theta(a) &= \frac{a(t_2 - t_1)}{\left[\int_0^{dd_1} e^{aD} dH(D) + e^{add_1}(1 - H(dd_1)) - \int_0^{dd_2} e^{aD} dH(D) - e^{add_2}(1 - H(dd_2))\right]} \\
&= \frac{t_2 - t_1}{\int_{dd_2}^{dd_1} e^{aD}(1 - H(D))dD}, \qquad (7)
\end{aligned}
$$

using integration by parts. We denote by θ^* and a^* the highest risk and risk aversion on this frontier.

The insurer chooses (t_1, dd_1, t_2, dd_2) by maximizing his expected profit. Similarly to (3), we have

$$
\begin{aligned}
\mathrm{E}[\pi] &= \sum_{c=1}^{2} \int_{\mathcal{A}_c} \left[t_c - \theta \int_{dd_c}^{\overline{d}} (1 - H(D))dD \right] dF(\theta, a) \\
&= \sum_{c=1}^{2} \nu_c \left[t_c - \mathrm{E}[\theta | \mathcal{A}_c] \int_{dd_c}^{\overline{d}} (1 - H(D))dD \right],
\end{aligned}
$$

where the second equality follows from $\int_{\mathcal{A}_c} \theta dF(\theta, a) = \nu_c \mathrm{E}[\theta|\mathcal{A}_c]$ with $\nu_c = \int_{\mathcal{A}_c} dF(\theta, a)$. The insurer's expected profit from selling the two coverages is a weighted average with weights ν_1 and ν_2 for the proportion of insurees choosing the first and second contracts, respectively.

The optimal contracts need also to satisfy insurees' incentive compatibility and participation constraints:

$$
\begin{aligned}
&CE(t_c, dd_c; \theta, a) > CE(t_{c'}, dd_{c'}, \theta, a), c \neq c', \quad \forall(\theta, a) \in \mathcal{A}_c, c = 1, 2, \\
&CE(t_c, dd_c; \theta, a) \geq CE(0, 0; \theta, a), \quad \forall(\theta, a) \in \mathcal{A}_c, c = 1, 2.
\end{aligned}
$$

Following Aryal and Perrigne (2009), the only constraint that binds is the individual rationality constraint for the $(\underline{\theta}, \underline{a})$ insuree, i.e. $CE(t_1, dd_1; \underline{\theta}, \underline{a}) = \overline{s}$. Maximizing $\mathrm{E}[\pi]$ with respect to (t_1, dd_1, t_2, dd_2) with respect to this participation constraint gives the first-order conditions

$$
\nu_1 + \int_{\underline{a}}^{a^*} \left[t_1 - \theta(a) \left\{ \int_{dd_1}^{\overline{d}} (1 - H(D))dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial t_1} da
$$

$$- \int_{a^*}^{\overline{a}} \left[t_2 - \theta(a) \left\{ \int_{dd_2}^{\overline{d}} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial t_1} da = \lambda \tag{8}$$

$$\int_{\underline{a}}^{a^*} \left[t_1 - \theta(a) \left\{ \int_{dd_1}^{\overline{d}} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial dd_1} da + \mathrm{E}[\theta | \mathcal{A}_1] \nu_1 (1 - H(dd_1))$$

$$- \int_{a^*}^{\overline{a}} \left[t_2 - \theta(a) \left\{ \int_{dd_2}^{\overline{d}} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial dd_1} da - \lambda \underline{\theta} e^{\underline{a} dd_1} (1 - H(dd_1)) = 0 \tag{9}$$

$$\int_{\underline{a}}^{a^*} \left[t_1 - \theta(a) \left\{ \int_{dd_1}^{\infty} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial t_2} da$$

$$+ \nu_2 - \int_{a^*}^{\overline{a}} \left[t_2 - \theta(a) \left\{ \int_{dd_2}^{\overline{d}} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial t_2} da = 0 \tag{10}$$

$$\int_{\underline{a}}^{a^*} \left[t_1 - \theta(a) \left\{ \int_{dd_1}^{\infty} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial dd_2} da + \mathrm{E}(\theta | \mathcal{A}_2) \nu_2 (1 - H(dd_2))$$

$$- \int_{a^*}^{\overline{a}} \left[t_2 - \theta(a) \left\{ \int_{dd_2}^{\overline{d}} (1 - H(D)) dD \right\} \right] f(\theta(a), a) \frac{\partial \theta(a)}{\partial dd_2} da = 0, \tag{11}$$

$$t_1 = \frac{\underline{\theta}}{\underline{a}} \left[\int_{dd_1}^{\overline{d}} \left(e^{\underline{a} D} - e^{\underline{a} dd_1} \right) dH(D) \right] \tag{12}$$

where λ is the Lagrangian multiplier associated with the participation constraint. See Aryal and Perrigne (2009) for the derivation of (8)–(12) reinterpreting their θ as the expected number of accidents.

3 Identification with a Continuum of Contracts

In this section, we consider the case in which a continuum of coverages is offered to each insuree. Though this is seldom the case in practice, this allows us to understand the problem of identification and the role played by the assumptions in identifying the model structure. The model structure is given by the joint distribution of risk and risk aversion $F(\cdot, \cdot)$ and the damage distribution $H(\cdot)$ given that the insuree's utility function is specified as CARA. Besides the specification of this utility function, the identification problem is nonparametric.[3] The problem of identification is to recover uniquely the structure

[3]The problem of identifying nonparametrically the agent's utility function is quite complex. In the context of auctions, the bidder's utility function is not identified in general. Nonparametric identification is achieved with the help of exclusion restrictions using exogenous variations in the number of bidders

$[F(\cdot,\cdot), H(\cdot)]$ from the observables. In the case of a continuum of contracts, we observe the contract purchased by each insuree (t, dd) and the J claims made by each insuree with the corresponding amounts of damages $(D_1, \ldots, D_J)$. In Section 3.2, we observe J^* claims with their corresponding damages $(D_1, \ldots, D_{J^*})$ because of the truncation at the deductible.

We introduce some observed variables characterizing the insuree. We distinguish two kinds of variables. The variables related to the insuree's personal information such as age, gender, education, marital status, location and driving experience are denoted by X, while the variables related to the insuree's car such as the car mileage, business use, car value, power, model and make are denoted by Z.[4] We remark that only X is an exogenous variable as Z can be viewed as endogenously determined in a model including the insuree's car choice, where Z becomes a funtion of (θ, a, X). In this section, we allow (θ, a) and (X, Z) to be dependent thereby allowing Z to be endogenous.

With the introduction of (X, Z) with values in the support $\mathcal{S}_{XZ} \subset I\!R^{\dim X + \dim Z}$, the model structure becomes $F(\theta, a|X, Z)$ and $H(D|X, Z)$ as we expect that both variables affect the insuree's risk and risk aversion and the damage. For instance, the amount of damage with an expensive car is likely to be larger than the damage with an inexpensive one. This intuition is supported by the empirical analysis of Cohen and Einav (2007) relying on some functional form for $F(\theta, a|X, Z)$. Let $G(\cdot|X, Z)$ denote the observed deductible distribution conditional on (X, Z). It is crucial that all the variables used by

as in Guerre, Perrigne and Vuong (2009) or with the help of additional data from ascending auction as in Lu and Perrigne (2008). See also Campo, Guerre, Perrigne and Vuong (2009) for semiparametric identification when the bidder's utility function is parameterized as CARA or CRRA. In the context of insurance with bidimensional screning, it is likely that the insurer's utility function is not identified. Moreover, the CARA specification simplifies considerably the derivation of the model through an explicit form of the certainty equivalence.

[4]The value of the car is used as a proxy for wealth w when computing the certainty equivalence so that w is a variable in Z. Given that only the value of the car is at risk in the case of an accident, we can consider that the relevant wealth in the model is the value of the car. Cohen and Einav (2007) use a different proxy for wealth obtained from additional census data on average income. This measure of wealth is then incorporated in the vector X in their empirical analysis.

the insurer to discriminate insurees are included in (X, Z).

In studies on identification of structural models, it is important to be precise about the set of admissible structures and the assumptions of the theoretical model. We formalize such assumptions made on the structure and (θ, a, J, D, X, Z). Specifically, the structure $[F(\cdot, \cdot|X, Z), H(\cdot|X, Z)]$ belongs to $\mathcal{F}_{XZ} \times \mathcal{H}_{XZ}$ as defined below.

Definition 1: *Let $\mathcal{F}_{XZ}$ be the set of conditional distributions $F(\cdot, \cdot|X, Z)$ satisfying*
(i) For every $(x, z) \in \mathcal{S}_{XZ}$, $F(\cdot, \cdot|x, z)$ is a c.d.f. with compact support $\Theta(x, z) \times \mathcal{A}(x, z) = [\underline{\theta}(x, z), \overline{\theta}(x, z)] \times [\underline{a}(x, z), \overline{a}(x, z)] \subset I\!R_+^* \times I\!R_+^*$,
(ii) The conditional density $f(\cdot, \cdot|\cdot, \cdot) > 0$ on its support.

Definition 2: *Let $\mathcal{H}_{XZ}$ be the set of distributions $H(\cdot|X, Z)$ satisfying*
(i) For every $(x, z) \in \mathcal{S}_{XZ}$, $H(\cdot|x, z)$ is a c.d.f with compact support $[0, \overline{d}(x, z)] \subset I\!R_+$ with $\sup_{(x,z)\in\mathcal{S}_{XZ}} \overline{d}(x, z) < +\infty$,
(ii) The conditional density $h(\cdot|\cdot, \cdot) > 0$ on its support.

Assumption 1: *We have*
(i) $(D_1, \ldots, D_J) \perp (\theta, a)|(J, X, Z)$.
(ii) $(D_1, \ldots, D_J)|(J, X, Z)$ are i.i.d. as $H(\cdot|X, Z)$,
(iii) $J \perp (X, Z, a)|\theta$ with $J|\theta \sim \mathcal{P}(\theta)$, i.e. $\Pr[J = j] = e^{-\theta}\frac{\theta^j}{j!}$.

Assumption 1-(i) says that conditional on the insuree's characteristics (X, Z), the amount of damage does not provide any information on his risk and risk aversion. For instance, conditional on (X, Z), the damage depends on factors such as road and weather conditions, bad luck which are independent of (θ, a). In the same spirit, Assumption 1-(ii) says that damages are independent conditional on (X, Z). Regarding Assumption 1-(iii), the number of accidents J depends on the insuree's risk θ only, while the Poisson distribution follows the theoretical model of Section 2, where the insuree's risk θ is the expected number of accidents. We maintain Assumption 1 throughout the paper. In addition, (θ, a, J, X, Z) is i.i.d. across insurees.

3.1 Case 1: Full Damage Distribution

Case 1 considers the best data scenario. In addition to a continuum of coverages offered to each insuree, the damage is observed for every accident whether its amount is below or above the deductible. It follows that $H(\cdot|X,Z)$ is identified on $[0,\overline{d}(X,Z)]$. It remains to study the identification of $F(\cdot,\cdot|X,Z)$. For the rest of Section 3, to simplify the notations, we suppress the conditioning on (X,Z). We first proceed by studying the identification of the distribution $K(\cdot)$ of certainty equivalence of no coverage in view of Section 2. If one can identify $K(\cdot)$, there is some hope to identify $F(\cdot,\cdot)$. The optimal contracts are characterized by (5) and (6). Equation (5) defines a one-to-one mapping between the certainty equivalence s and the deductible dd, while (6) defines a one-to-one mapping between dd and t. The key idea is to exploit the former mapping to identify the distribution of certainty equivalence from the observed deductible distribution $G(\cdot)$. This result is in the spirit of the nonparametric identification literature on auctions and contracts.[5] We have $G(dd) = \Pr(\tilde{dd} \le dd) = \Pr(\tilde{s} \le s(dd)) = K(s)$ implying $g(dd) = k(s)s'(dd)$ with $s(\cdot)$ the inverse of $dd(\cdot)$ by monotonicity of the latter. Hence,

$$\frac{G(dd)}{g(dd)} = \frac{K(s)}{k(s)}\frac{1}{s'(dd)} = \frac{K(s)}{k(s)}dd'(s).$$

Substituting the above expression in (5), we obtain

$$\eta(s,a(s),dd(s))\mathrm{E}[\theta|s](1-H(dd)) + \frac{G(dd)}{g(dd)}\left\{-\frac{\frac{\partial\eta(s,a(s),dd(s))}{\partial dd}}{\eta(s,a(s),dd(s))} + \frac{\eta'(s,a(s),dd(s))}{\eta(s,a(s),dd(s))}s'(dd)\right\} = 1.$$

From (6), we have $t'_+(dd) = -1/\eta(s,a(s),dd(s))$, where $t_+(dd) = t[s(dd)]$. We also have $dt'_+(dd(s))/ds = -d[\eta(s,a(s),dd(s))]^{-1}/ds$, i.e. $t''_+(dd) \times dd'(s) = \eta'(s,a(s),dd(s))/[\eta(s,$

[5] For auctions, see Guerre, Perrigne and Vuong (2000) and the survey by Athey and Haile (2007) where the mapping between the observed bid and the unobserved private value is exploited to identify the private value distribution. For contracts, see Perrigne and Vuong (2009) in the context of nonlinear pricing, and Perrigne and Vuong (2008) in the context of a procurement model with adverse selection and moral hazard. There, the mapping between the observed price or quantity and the unobserved firm's type or efficiency is exploited to recover the underlying distribution of firms' efficiency and willingness to pay, respectively. See also D'Haultfoeuille and Fevrier (2007).

$a(s), dd(s))]^2$ or equivalently $t''_+(dd) = [\eta'(s, a(s), dd(s))/[\eta(s, a(s), dd(s))]^2] \times s'(dd)$. Using this result, we can rewrite the previous equation as

$$\mathrm{E}(\theta|s)(1 - H(dd)) + \frac{G(dd)}{g(dd)} \left\{ -\frac{\frac{\partial \eta(s,a(s),dd(s))}{\partial dd}}{\eta(s, a(s), dd(s))^2} + t''_+(dd) \right\} = -t'_+(dd).$$

From (4), the derivative of $\eta(\cdot, \cdot, \cdot)$ with respect to dd is

$$\frac{\partial \eta(s, a(s), dd(s))}{\partial dd} = -\eta(s, a(s), dd(s)) \left[a(s) - \frac{h(dd)}{1 - H(dd)} \right].$$

Thus, the first-order condition defining the optimal deductible can be rewritten as

$$\mathrm{E}(\theta|dd)(1 - H(dd)) + \frac{G(dd)}{g(dd)} \left[-t'_+(dd) \left(a(s) - \frac{h(dd)}{1 - H(dd)} \right) + t''_+(dd) \right] = -t'_+(dd),$$

where $\mathrm{E}(\theta|s) = \mathrm{E}(\theta|dd)$ because of the one-to-one mapping between dd and s. After elementary algebra, we obtain

$$a(s) = \frac{1}{t'_+(dd(s))} \left\{ \frac{g(dd)}{G(dd)} \left[t'_+(dd(s)) + \mathrm{E}(\theta|dd)(1 - H(dd)) \right] + t''_+(dd(s)) \right\} + \frac{h(dd)}{1 - H(dd)},$$

showing that $a(s)$ is identified as the right-hand side is observed or identified from observables. In particular, $\mathrm{E}(\theta|dd)$ is identified by the expected number of claims made by insurees choosing the deductible dd given that all the claims including those below the deductible are observed by assumption, i.e. $\mathrm{E}(\theta|dd) = \mathrm{E}(J|dd)$.[6] But, using (4) we have

$$s = w + \frac{t'(dd)(\phi_a - 1)}{a(s)\exp(a(s)dd)(1 - H(dd))},$$

showing that s is identified from the knowledge of dd. Thus, we have the following result.

Lemma 1: *Suppose that a continuum of insurance coverages is offered to each insuree and all accidents are observed for each insuree. Under Assumption 1, the pair* $(K(\cdot), H(\cdot))$ *is identified.*

It remains to investigate whether we can identify $F(\cdot, \cdot)$ from the knowledge of $K(\cdot)$. A sketch of the argument is as follows, where the observed number of claims J plays a

[6]We have $\mathrm{E}[J|dd] = \mathrm{E}[J|s] = \mathrm{E}\{\mathrm{E}[J|\theta, s]|s\} = \mathrm{E}\{\mathrm{E}[J|\theta, a]|s\} = \mathrm{E}\{\mathrm{E}[J|\theta]|s\} = \mathrm{E}[\theta|s]$, where we have used Assumption 1-(iii) and the one-to-one mapping between (θ, a) and (θ, s).

crucial role in identifying $F(\cdot,\cdot)$.[7] Specifically, from the moment generating function of the number of accidents J conditional on s, we show that we can identify the moment generating function of θ given s in a neighborhood of zero. As is well known, the latter identifies $F_{\theta|S}(\cdot|\cdot)$. Once we identify $F_{\theta|S}(\cdot|\cdot)$, we use $K(\cdot)$ to derive the joint distribution of (θ, s). Identification of the joint density of (θ, a) follows from the known one-to-one mapping between (θ, s) and (θ, a) given by (1).

Formally, for a given certainty equivalence s, the subpopulation of insurees with insurance coverage $(t(s), dd(s))$ and their corresponding claims gives the moment generating function $M_{J|S}(\cdot|s)$ as

$$\begin{aligned} M_{J|S}(t|s) &= \mathrm{E}[e^{Jt}|S=s] = \mathrm{E}\left\{\mathrm{E}[e^{Jt}|\theta,S]|S=s\right\} \\ &= \mathrm{E}\left\{\mathrm{E}[e^{Jt}|\theta,a]|S=s\right\} = \mathrm{E}\left\{\mathrm{E}[e^{Jt}|\theta]|S=s\right\} \\ &= \mathrm{E}\left\{e^{\theta(e^t-1)}|S=s\right\} = M_{\theta|S}(e^t-1|s), \end{aligned} \tag{13}$$

where the third equality follows from the one-to-one mapping between (θ, s) and (θ, a) and the fourth and fifth equalities from Assumption 1-(iii) using the moment generating function of the Poisson distribution with parameter θ. In particular, the above equation shows that the moment generating function $M_{J|S}(\cdot|s)$ exists for every $t \in I\!R$ because θ has a compact support given $S = s$. Moreover, letting $u = e^t - 1$ shows that

$$M_{\theta|S}(u|s) = M_{J|S}(\log(1+u)|s)$$

for all $u \in (-1, +\infty)$. Thus $M_{\theta|S}(\cdot|s)$ is identified on a neighborhood of 0 thereby identifying $F_{\theta|S}(\cdot|s)$. See (say) Billingsley (1995, p. 390).[8]

The joint density of (θ, s) is $f(\theta, s) = f(\theta|s)k(s)$, which is identified. From the known one-to-one mapping $T(\cdot,\cdot)$ that transforms (θ, a) into (θ, s), namely $T(\theta, a) = [\theta, w -$

[7] In contrast, if the analyst observes only whether $J = 0$ or $J \geq 1$ with the risk measured by the probability of some accident(s) $\tilde{\theta} = 1 - e^{-\theta}$, $F(\cdot,\cdot)$ is not identified as shown by Aryal, Perrigne and Vuong (2009).

[8] Alternatively, because $M_{\theta|S}(\cdot|s)$ exists in a neighborhood of 0, then all the moments of θ given $S = s$ are identified by $M^{(k)}_{\theta|S}(0|s) = \mathrm{E}[\theta^k|S=s]$ for $k = 0, 1 \ldots$. Since θ given s has compact support, we are in the class of Hausdorff moment problems, which are always determinate, i.e. the distribution of θ given s is uniquely determined by its moments.

$[\theta(\phi_a - 1)]/a]'$ with $\phi_a = \int e^{aD} dH(D)$ and $H(\cdot)$ known, we can recover the joint distribution of (θ, a) as

$$f(\theta, a) = f_{\theta S}(T^{-1}(\theta, a)) \left| \frac{\partial T^{-1}(\theta, a)}{\partial(\theta, a)} \right|.$$

This result is formally stated in the following proposition.

Proposition 1: *Suppose that a continuum of insurance coverages is offered to each insuree and all accidents are observed for each insuree. Under Assumption 1, the structure $(F(\cdot,\cdot), H(\cdot))$ is identified.*

3.2 Case 2: Truncated Damage Distribution

We maintain the assumption that the insurer offers a continuum of contracts to each insuree but we now consider that the damage distribution is not fully observed. In practice and making abstraction of dynamic considerations, an accident leads to a claim if and only if the damage is above the deductible. Using the claim data only, we cannot identify the damage distribution but only the truncated damage distribution on $[dd, \overline{d}]$. Nonetheless, the damage distribution is still identified on its support $[0, \overline{d}]$ by exploiting claim data for insurees buying full insurance for whom the deductible is zero. For this coverage, every accident is reported since $dd = 0$ and thus $H(\cdot)$ is identified. Specifically, $H_{D|dd}(\cdot|0) = H_{D|S}(\cdot|\underline{s}) = H_{D|(\theta,a)}(\cdot|\overline{\theta}, \overline{a}) = H_D(\cdot)$ by Assumption 1-(i). Thus, we have the following lemma.

Lemma 2: *Under Assumption 1, $H(\cdot)$ is identified.*

It remains to study the identification of $F(\cdot,\cdot)$. Though the reported number of accidents J^* is observed instead of J, the argument is similar to Case 1. Specifically, reviewing the argument leading to Lemma 1, it is straightforward to see that $K(\cdot)$ is identified if $\mathrm{E}(\theta|dd)$ is. Since accidents are reported only if the damage is above the deductible, we have $\mathrm{E}[\theta|dd] \neq \mathrm{E}[J^*|dd]$, where J^* is the number of reported accidents, i.e. those with a damage above the deductible. But J^* given (J, dd) is distributed as a Binomial with parameters $(J, 1 - H(dd))$ by Assumption 1-(i,ii). Thus, $\mathrm{E}[J^*|dd] =$

$\mathrm{E}\{\mathrm{E}[J^*|J,dd]|dd\} = \mathrm{E}[J(1-H(dd))|dd] = (1-H(dd))\mathrm{E}[J|dd] = (1-H(dd))\mathrm{E}(\theta|dd)$, i.e. $\mathrm{E}[\theta|dd] = \mathrm{E}[J^*|dd]/(1-H(dd))$. Hence, $\mathrm{E}[\theta|dd]$ is identified despite the truncation of the damage distribution at dd leading to the identification of $K(\cdot)$.

Turning to the identification of $F(\theta,a)$, we begin with the identification of $F_{\theta|S}(\cdot|\cdot)$ as before. The moment generating function of J^* given s is

$$\begin{aligned} M_{J^*|S}(t|s) &= \mathrm{E}[e^{J^*t}|S=s] = \mathrm{E}\{\mathrm{E}[e^{J^*t}|J,S]|S=s\} = \mathrm{E}\{\mathrm{E}[e^{J^*t}|J,dd]|S=s\} \\ &= \mathrm{E}\left\{[H(dd)+(1-H(dd))e^t]^J|S=s\right\} = \mathrm{E}\left\{e^{J\log[H(dd)+(1-H(dd))e^t]}|S=s\right\} \\ &= M_{\theta|S}\left[e^{\log[H(dd)+(1-H(dd))e^t]}-1|s\right] = M_{\theta|S}[(1-H(dd))(e^t-1)|s] \qquad (14) \end{aligned}$$

where the fourth equality uses the moment generating function of the Binomial distribution $\mathcal{B}(J,1-H(dd))$ and the fifth equality uses (13) with t replaced by $\log[H(dd)+(1-H(dd))e^t]$. Thus, we obtain

$$M_{\theta|S}(u|s) = M_{J^*|S}\left[\log\left(1+\frac{u}{1-H(dd)}\right)|s\right],$$

for $u \in (-(1-H(dd),+\infty)$. The rest of the argument in Case 1 applies leading to the following proposition.

Proposition 2: *Suppose that a continuum of insurance coverages is offered to each insuree and accidents are observed if and only if the damage is above the deductible. Under Assumption 1, the structure $(F(\cdot,\cdot),H(\cdot))$ is identified.*

4 Identification with a Finite Number of Contracts

We now address identification of the model when only (say) two contracts are offered given (X,Z). The identification argument can no longer rely on the identification of the density of certainty equivalence as we cannot exploit the one-to-one mapping between the insuree's certainty equivalence and his deductible choice. There is a continuum of $s \in [\underline{s},\overline{s}]$ values, while there are only a finite number of deductibles. Consequently, the FOCs characterizing (t_1,dd_1,t_2,dd_2) alone will not allow us to identify $F(\theta,a)$. In addition to the key role played by the observed number of claims, we exploit sufficient variations in exogenous variables to achieve identification. In particular, the optimality of contracts is used through the contract form and the screening procedure. As before, we distinguish whether the full damage or truncated damage distribution is observed.

4.1 Case 3: Full Damage Distribution

This case is the closest to Cohen and Einav (2007) who consider that claim data contain all the accidents. Cohen and Einav (2007) identify the joint distribution of risk and risk aversion under parametric assumptions. In particular, they assume a lognormal mixture of Poisson for the claim data. Moreover, they do not exploit any information provided by the optimality of contracts. In this section, we show how some features of contract optimality combined with a full support assumption with sufficient variations in the car characteristics can be exploited to identifiy nonparametrically $f(\theta, a)$. In view of Cohen and Einav (2007) empirical findings, our identification result is important for several reasons. First, the nonparametric identification of the joint distribution of risk and risk aversion offers more flexibility on the dependence between risk and risk aversion. Their empirical findings display a counterintuitive positive correlation between the latter while one could expect a negative one. Second, their robustness analysis suggests that the offered contracts are suboptimal with their estimated negative correlation, i.e. the insurer could increase his profit by adjusting upward the current low deductibles. On the other hand, a positive correlation would imply lower levels of deductibles.

Our identification results rely on a nonparametric mixture of Poisson distribution for the number of claims. Specifically, the probability of the observed claims J conditional on some characteristics (x, z) is given by

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

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

Given that all the accidents and their corresponding damages are observed, the damage distribution $H(\cdot|X, Z)$ is identified. To establish identification of $F(\theta, a|X, Z)$, we proceed as follows. We first show the identification of the marginal distribution of θ given (X, Z)

following an argument similar to Case 1. In a second step, we identify the conditional distribution of a given (θ, X, Z) at $a(\theta, X, Z)$, which defines the frontier between the two sets $\mathcal{A}_1(X,Z)$ and $\mathcal{A}_2(X,Z)$. In a third step, we make an exclusion restriction and a full support assumption involving the car characteristics Z to achieve identification of the distribution of a given (θ, X, Z) on its support.

For the first step, we exploit again the observed number of accidents. Using an argument similar to that leading to (13) for the subpopulation of insurees with characteristics (x, z), the moment generating function $M_{J|X,Z}(\cdot|x,z)$ is

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

where the third and fourth equalities follow from Assumption 1-(iv). Thus, $f_{\theta|X,Z}(\cdot|\cdot,\cdot)$ is identified by its moment generating function

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

for all $u \in (-1, +\infty)$.

In the second step, we consider the probability that an insuree with risk θ and characteristics (X, Z) chooses the coverage $(t_1(X,Z), dd_1(X,Z))$ as intuitively this provides information about the insuree's risk aversion a. To do so, we define the discrete variable χ, which takes the values 1 and 2 whether the insuree chooses the coverage $(t_1(X,Z), dd_1(X,Z))$ or $(t_2(X,Z), dd_2(X,Z))$, i.e. whether the insuree's types (θ, a) belongs to $\mathcal{A}_1(X,Z)$ or $\mathcal{A}_2(X,Z)$, respectively. Thus, $\chi = 1$ is also equivalent to $a \leq a(\theta, X, Z)$, where the latter is the inverse of the frontier (7), where (t_1, dd_1, t_2, dd_2) and $H(\cdot)$ now depends on (X, Z). We remark that some features of optimal contracts are used here, namely the offered contracts are of the form premium/deductible, while the (θ, a) space is partitioned optimally by the frontier (7). The above probability of interest can then be written as $\Pr[\chi = 1|\theta, X = x, Z = z]$, which is

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

by Bayes rule, where $\nu_1(x,z)$ is the proportion of insurees with characteristics (x,z) choosing the coverage $(t_1(x,z), dd_1(x,z))$. The latter is identified from the data. Since $f_{\theta|X,Z}(\cdot|\cdot,\cdot)$ is identified from the first step, it remains to identify $f_{\theta|\chi,X,Z}(\cdot|1,x,z)$. Applying the same argument as in Step 1 but conditioning on $\chi = 1$ as well, we obtain

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

where the second equality follows from that conditioning on (θ, a, χ) is equivalent to conditioning on (θ, a), while the third equality follows as before from Assumption 1-(iv). Thus, $f_{\theta|\chi,X,Z}(\cdot|1,\cdot,\cdot)$ is identified by its moment generating function

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

for all $u \in (-1, +\infty)$.[9] Hence, $F_{a|\theta,X,Z}[a(\theta,x,z)|\theta,x,z]$ is identified for every $\theta \in [\underline{\theta}(x,z), \overline{\theta}(x,z)]$ and $(x,z) \in \mathcal{S}_{XZ}$.

To conduct policy counterfactuals, however, the analyst may need to identify $F(\cdot,\cdot|x,z)$ on the whole support $\Theta(x,z) \times \mathcal{A}(x,z)$. This is the purpose of the third step. To do so, we make the following assumptions. Let $\mathcal{S}_W$ denote the support of some variable W and $\mathcal{S}_{W_1|w2}$ denote the support of some variable W_1 given some variable $W_2 = w_2$.

Assumption 2: *We have*
(i) $a \perp Z|(\theta, X)$
(ii) $\forall(\theta,a,x) \in \mathcal{S}_{\theta a X}$*, there exists* $z \in \mathcal{S}_{Z|\theta x}$ *such that* $a(\theta,x,z) = a$.

Assumption 2-(i) is an exclusion restriction that gives

$$F_{a|\theta,X,Z}(a(\theta,x,z)|\theta,x,z) = F_{a|\theta,X}(a(\theta,x,z)|\theta,x) \quad \forall(\theta,x,z).$$

Because the left-hand side is identified from the second step, sufficient variations in $a(\theta,x,z)$ due to z can identify $F_{a|\theta,X}(\cdot|\theta,x)$. This is the purpose of Assumption 2-(ii), which is a full support assumption. Similar assumptions (sometimes called large support

[9] The argument works as well by considering $\chi = 2$ leading to the overidentification of $F_{a|\theta,X,Z}[a(\theta,x,z)|\theta,x,z]$. This issue will be further discussed in Section 5.

assumptions) have been made by different authors in various contexts. See Matzkin (1992, 1993), Lewbel (2000), Carneiro, Hansen and Heckman (2003), Berry and Haile (2009) and Imbens and Newey (2009) among others. In our context, this assumption can be interpreted as follows: For every individual with personal characteristics (θ, a, X), there exists some (say) car characteristics Z for which the insuree is indifferent between the two offered coverages. The full support assumption is sufficient to guarantee identification as shown next but it is not necessary.[10] Specifically, we have

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

where the first equality uses the full support assumption and the second equality uses the exclusion restriction assumption. Note that $a(\cdot, \cdot, \cdot)$ is identified in view of (7). Identification of $F(\theta, a|x, z)$ follows using the first step. This result is formally stated in the next proposition.

Proposition 3: *Suppose that two insurance coverages are offered to each insuree and all accidents are observed for each insuree. Under Assumptions 1 and 2, the structure* $(F(\cdot, \cdot|X, Z), H(\cdot|X, Z))$ *is identified.*

4.2 Case 4: Truncated Damage Distribution

The data scenario analyzed in Case 4 corresponds to the insurance data that a researcher typically has, i.e. a finite number of contracts offered with claims filed only if damages are above the deductible. Case 3 has shown that observing a finite number of contracts does not prevent the nonparametric identification of the joint distribution of risk and risk aversion provided all accident information is available and there is enough variation in some excluded exogenous variables. In contrasts, the truncation on the damage distribution in Case 4 will limit the extent of identification. Nevertheless, we show that $F(\cdot, \cdot)$ is

[10] In other words, it may not be the minimal assumption required to identify $F(\theta, a|x, z)$. For instance, we could have switched the roles of X and Z in Assumption 2. We prefer to use Z because it contains the car value, which is continuous as required by the full support assumption.

identified up to the knowledge of $H(dd_2(X,Z)|X,Z)$ or equivalently $H(dd_1(X,Z)|X,Z)$, where $dd_1(X,Z) > dd_2(X,Z)$.[11]

We follow similar steps as in Case 3 with $\tilde{\theta} \equiv (1 - H(dd_2(X,Z)|X,Z))\theta$ replacing θ while modifying the argument as J is unobserved. To begin, we note the relationship between $1 - H(dd_1(X,Z)|X,Z)$ and $1 - H(dd_2(X,Z)|X,Z)$ which allows us to focus on identification only in terms of $1 - H(dd_2(X,Z)|X,Z)$. Because a claim is filed only if it involves a damage above the deductible, we identify the truncated damage distributions

$$H_c^*(\cdot|X,Z) = \frac{H(\cdot|X,Z) - H(dd_c(X,Z)|X,Z)}{1 - H(dd_c(X,Z)|X,Z)},$$

on $[dd_c(X,Z), \overline{d}(X,Z)]$ from the subpopulation of insurees buying the coverage $(t_c(X,Z), dd_c(X,Z))$ for $c = 1,2$. To simplify the notations, we let $H_c(X,Z) = H(dd_c(X,Z)|X,Z)$ hereafter. Differentiating the above equations and taking their ratio show that

$$\pi(X,Z) \equiv \frac{h_2^*(D|X,Z)}{h_1^*(D|X,Z)} = \frac{1 - H_1(X,Z)}{1 - H_2(X,Z)}, \tag{15}$$

for all $D \geq dd_1(X,Z)$, where $0 < \pi(X,Z) < 1$. In particular, the function $\pi(\cdot,\cdot)$ is identified from the data, while $H(\cdot|X,Z)$ is identified on $[dd_2(X,Z), \overline{d}(X,Z)]$ up to the knowledge of $H_2(X,Z)$.

To identify the marginal density $\tilde{f}_{\tilde{\theta}|XZ}(\cdot|\cdot,\cdot)$ of $\tilde{\theta}$ given (X,Z), we exploit the observed number of reported accidents J_c^*. Using a similar argument as in (14), the moment generating function of J^* given (χ, X, Z) is

$$\begin{aligned}
M_{J^*|\chi,X,Z}(t|c,x,z) &= \mathrm{E}[e^{J^*t}|\chi = c, X = x, Z = z] \\
&= \mathrm{E}\{\mathrm{E}[e^{J^*t}|J,\chi,X,Z]|\chi = c, X = x, Z = z\} \\
&= \mathrm{E}\left\{[H_\chi(X,Z) + (1 - H_\chi(X,Z))e^t]^J|\chi = c, X = x, Z = z\right\} \\
&= \mathrm{E}\left\{\mathrm{E}[e^{J\log[H_\chi(X,Z)+(1-H_\chi(X,Z))e^t]}|\theta,\chi,X,Z]|\chi = c, X = x, Z = z\right\} \\
&= \mathrm{E}\left[e^{\theta[H_\chi(X,Z)+(1-H_\chi(X,Z))e^t-1]}|\chi = c, X = x, Z = z\right] \\
&= M_{\theta|\chi,X,Z}[(1 - H_\chi(X,Z))(e^t - 1)|c,x,z] \qquad (16)
\end{aligned}$$

[11] When two contracts are offered, it is never optimal for the insurer to offer full insurance, i.e. $dd_2(X,Z) = 0$. Therefore, we cannot use the argument of Case 2 to identify $H(\cdot|X,Z)$ and hence $H_2(\cdot|X,Z)$.

where the third equality uses the moment generating function of J^* given χ, X, Z, which is distributed as a Binomial $\mathcal{B}(J, 1-H_\chi(X,Z))$ by Assumption 1-(ii,iii), and the fifth equality follows from Assumption 1-(iv) and the moment generating function of the Poisson distribution. Thus, we obtain

$$M_{\theta|\chi,X,Z}[u|c,x,z] = M_{J^*|\chi,X,Z}\left[\log\left(1+\frac{u}{1-H_\chi(X,Z)}\right)|c,x,z\right], \tag{17}$$

for $u \in (-(1-H_\chi(X,Z),+\infty)$. In particular, the distribution of risk θ given (χ, X, Z) is identified up to the knowledge of $H_\chi(X,Z)$. Since $\tilde{\theta} = (1-H_2(X,Z))\theta$, its moment generating function given (χ, X, Z) is

$$\begin{aligned} M_{\tilde{\theta}|\chi,X,Z}(u|c,x,z) &= M_{\theta|\chi,X,Z}(u(1-H_2(x,z))|c,x,z) \\ &= \begin{cases} M_{J^*|\chi,X,Z}\left[\log\left(1+\frac{u}{\pi(x,z)}\right)|1,x,z\right] & \text{if } c=1, \\ M_{J^*|\chi,X,Z}\left[\log\left(1+u\right)|2,x,z\right] & \text{if } c=2, \end{cases} \end{aligned} \tag{18}$$

for all $u \in (-\pi(x,z),+\infty)$ and $u \in (-1,+\infty)$, respectively. Thus, the moment generating function of $\tilde{\theta}$ given (X,Z) is

$$\begin{aligned} M_{\tilde{\theta}|X,Z}(u|x,z) &= \mathrm{E}\{\mathrm{E}[e^{u\tilde{\theta}}|\chi,X,Z]|X=x,Z=z\} \\ &= M_{J^*|\chi,X,Z}\left[\log\left(1+\frac{u}{\pi(x,z)}\right)|1,x,z\right]\nu_1(x,z) \\ &\quad + M_{J^*|\chi,X,Z}\left[\log\left(1+u\right)|2,x,z\right]\nu_2(x,z), \end{aligned} \tag{19}$$

for $u \in (-\pi(x,z),+\infty)$, showing that $\tilde{f}_{\tilde{\theta}|X,Z}(\cdot|\cdot,\cdot)$ is identified as $\pi(X,Z)$, $\nu_1(X,Z)$ and $\nu_2(X,Z)$ are known from the data. Since $f_{\theta|X,Z}(\theta|x,z) = (1-H_2(x,z))\tilde{f}_{\tilde{\theta}|X,Z}((1-H_2(x,z))\theta|X,Z)$, the former density is identified up to $H_2(x,z)$.

In the second step, we consider again the probability that an insuree with risk θ and characteristics (X,Z) chooses the coverage $(t_1(X,Z), dd_1(X,Z))$. Using (7) and $1 - H(D|X,Z) = (1-H_2(X,Z))(1-H_2^*(D|X,Z))$, we remark that the optimal frontier between buying the two coverages in the space $(\tilde{\theta}, a)$ is given by

$$\tilde{\theta}(a,X,Z) = \frac{t_2(X,Z)-t_1(X,Z)}{\int_{dd_2(X,Z)}^{dd_1(X,Z)} e^{aD}[1-H_2^*(D|X,Z)]dD}, \tag{20}$$

leading to the inverse $\tilde{a}(\tilde{\theta}, X, Z)$, which is identified. As before, from Bayes rule we have

$$F_{a|\tilde{\theta},X,Z}(\tilde{a}(\tilde{\theta}, x, z)|\tilde{\theta}, x, z) = \frac{\tilde{f}_{\tilde{\theta}|\chi,X,Z}(\tilde{\theta}|1, x, z)\nu_1(x, z)}{\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x, z)}, \tag{21}$$

where $\nu_1(x, z)$ and $\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x, z)$ are identified. Moreover, $\tilde{f}_{\tilde{\theta}|\chi,X,Z}(\cdot|1, x, z)$ is identified because its moment generating function $M_{\tilde{\theta}|\chi,X,Z}(\cdot|1, x, z)$ is identified on $(-\pi(x, z), +\infty)$ as seen above.

In the third step, we note that $F_{a|\tilde{\theta},X,Z}(\tilde{a}(\tilde{\theta}, x, z)|\tilde{\theta}, x, z) = F_{a|\theta,X,Z}(a(\theta, x, z)|\theta, x, z)$ thereby identifying the latter up to $H_2(x, z)$. The rest of the argument is exactly the same as in Case 3 leading to the identification of $F_{a|\theta,X}(\cdot|\cdot, \cdot)$ and then the joint distribution of (θ, a) given (X, Z) up to the knowledge of $H_2(X, Z)$ because $a(\cdot, \cdot, \cdot)$ is known up to $H_2(\cdot, \cdot)$. We have then proved the following result.

Proposition 4: *Suppose that two insurance coverages are offered to each insuree and accidents are observed only when damages are above the deductible. Under Assumptions 1 and 2, the structure $(F(\cdot, \cdot|X, Z), H(\cdot|X, Z))$ is identified up to $H_2(X, Z)$.*

Up to now, we have used little of the optimality of the offered coverages beyond the contract form and the screening through the optimal frontier partitioning the insurees' types. For instance, we have not used the FOC (8)–(12) determining the optimal insurance terms $(t_1(X, Z), dd_1(X, Z), t_2(X, Z), dd_2(X, Z))$. One might ask whether the use of these FOC may help in identifying some features of the structure or even the full structure itself. For instance, we note that (12) identifies $\underline{a}(X, Z)$ because the latter solves the identifying equation

$$t_1(X, Z) = \frac{\tilde{\underline{\theta}}(X, Z)}{\underline{a}(X, Z)} \int_{dd_1(X,Z)}^{\overline{d}(X,Z)} \left(e^{\underline{a}(X,Z)D} - e^{\underline{a}(X,Z)dd_1(X,Z)}\right) h_2^*(D|X, Z)dD,$$

using $h(D|X, Z) = [1-H_2(X, Z)]h_2^*(D|X, Z)$ and $\tilde{\underline{\theta}}(X, Z) = \underline{\theta}(X, Z)[1-H_2(X, Z)]$. Other features of the structure may be identified.

A consequence of Proposition 4 is that the structure $[F(\cdot, \cdot|X, Z), H(\cdot|X, Z)]$ is identified if and only if $H_2(X, Z)$ is identified. The next lemma shows that $H_2(X, Z)$ is not identified even when considering the full optimality of the model including the FOC (8)-(12).[12]

Lemma 3: *Suppose that two insurance coverages are offered to each insuree and accidents are observed only when damages are above the deductible. Under Assumptions 1 and 2,* $H_2(X,Z)$ *is not identified.*

The proof can be found in the appendix. It relies on the construction of an observationally equivalent structure leading to the same observations. The nonidentification arises from a compensation between the increase in the number of accidents and an appropriate decrease in the probability of damages being greater than the deductible. From the insuree's perspective, such a compensation maintains the relative ranking between the two contracts. Thus, if a (θ, a) insuree buys $(t_1(X,Z), dd_1(X,Z))$ then the $((1-H_2(X,Z))\theta, a)$ insuree also buys the same coverage if there is an appropriate increase (decrease) in the probability of damages being greater than $dd_1(X,Z)$ thereby increasing (decreasing) the likelihood of getting indemnity from the insurer. For the insurer's perspective, the decrease (increase) in the average number of accidents is compensated by an appropriate decrease (increase) in the probability that the damage is below the deductible. Thus the expected payment to the insuree remains the same under either coverage.

5 Model Restrictions

This section derives the restrictions imposed by the model on observables under the data scenario of Case 4, i.e. a finite number of contracts and a truncated damage distribution. These restrictions can be used to test the model validity. For every insuree, we observe $[J^*, D_1^*, \ldots, D_{J^*}^*, \chi, T, DD, X, Z]$, where D_j^* denotes the (truncated) damage for the jth reported accident and (T, DD) are the premium and deductible chosen by the insuree. From the model, T and DD are given by $T = t_\chi(X,Z)$ aand $DD = dd_\chi(X,Z)$, where

[12] In particular, the observed optimal proportion of insurees $\nu_2(X,Z)$ does not help in identifying $H_2(X,Z)$. Specifically, $\nu_2(X,Z) = \int 1\!\!1[\theta \geq \theta(a,X,Z)] f(\theta,a) d\theta da = 1 - \int F_{a|\theta,X,Z}[a(\theta,X,Z)|\theta,X,Z] f_{\theta|X,Z}(\theta|X,Z) d\theta = 1 - \nu_1(X,Z)$, which is always true.

$t_c(X,Z)$ and $dd_c(X,Z)$ for $c=1,2$ are deterministic functions of (X,Z) satisfying the first-order conditions (8)-(12). Thus, the vector of observables has a joint distribution $\Psi(\cdot,\ldots,\cdot)$ with a density $\psi(\cdot,\ldots,\cdot) = \psi_{D_1^*,\ldots,D_{J^*}^*|J^*,\chi,X,Z}(\cdot,\ldots,\cdot|\cdot,\ldots,\cdot)\times\psi_{J^*|\chi,X,Z}(\cdot|\cdot,\cdot\cdot)\times \psi_{\chi|X,Z}(\cdot|\cdot,\cdot)\times\psi_{X,Z}(\cdot,\cdot)$.

The next lemma provides necessary and sufficient conditions on the joint distribution $\Psi(\cdot,\ldots,\cdot)$ to be rationalized by a structure $[F(\cdot,\cdot|\cdot,\cdot),H(\cdot|\cdot,\cdot)]\in\mathcal{F}_{XZ}\times\mathcal{H}_{XZ}$. We introduce some notations. Let $\mathcal{H}^*_{cXZ}$ be defined as the set $\mathcal{H}_{XZ}$ with the difference that the support is $[dd_c(X,Z),\overline{d}(X,Z)]$ for $c=1,2$. The remaining notations are introduced to write the model restrictions implied by the full support assumption and the first-order conditions (8)–(12) (see conditions (iv) and (v) below). The insurer's expected payment in case of accident given the coverage c and characteristics (x,z) is denoted $\mathrm{E}[P|c,x,z]=\int_{dd_c(x,z)}^{\overline{d}(x,z)}(1-\Psi_{D^*|\chi,X,Z}(D|c,x,z))dD$ for $c=1,2$. We define $\tilde\theta(a)\equiv\tilde\theta(a,x,z)$ and $\tilde a(\theta)\equiv\tilde\theta^{-1}(\tilde\theta,x,z)$ as in (20) with $H_2^*(D|X,Z)$ replaced by $\Psi_{D^*|\chi,X,Z}(D|2,X,Z)$. Let $\tilde f_{\tilde\theta|\chi,X,Z}(\cdot|\cdot,\cdot,\cdot)$ and $\tilde f_{\tilde\theta|X,Z}(\cdot|\cdot,\cdot)$ be densities corresponding to the moment generating functions (18) and (19), respectively with $\nu_c(x,z)$ replaced by $\psi_{\chi|X,Z}(c|x,z)$ for $c=1,2$ and $\pi(x,z)=\psi_{D^*|\chi,X,Z}(\cdot|2,x,z)/\psi_{D^*|\chi,X,Z}(\cdot|1,x,z)$. We denote by $\underline{\tilde\theta}=\underline{\tilde\theta}(x,z)$ the lower bound of the support of $\tilde f_{\tilde\theta|X,Z}(\cdot|\cdot,\cdot)$. Let $\tilde f(\cdot,\cdot|\cdot,\cdot)=\tilde f_{a|\tilde\theta,X,Z}(\cdot|\cdot,\cdot,\cdot)\tilde f_{\tilde\theta|X,Z}(\cdot|\cdot,\cdot)$, where $\tilde f_{a|\tilde\theta,X,Z}(\cdot|\cdot,\cdot,\cdot)$ is obtained from (21) and Assumption 2. Let $[\underline{a},\overline{a}]\equiv[\underline{a}(x,z),\overline{a}(x,z)]$ be the support of $\tilde f_{a|X,Z}(\cdot|x,z)$, while $a^*\equiv a^*(x,z)=\min\{\overline{a},\tilde a(\underline{\tilde\theta},x,z)\}$. Lastly, we define

$$\begin{aligned}\lambda(x,z) &= \psi_{\chi,X,Z}(1,x,z)+\int_{\underline{a}}^{a^*}\Big[t_1(x,z)-\tilde\theta(a)\mathrm{E}[P|1,x,z]\Big]\tilde f(\tilde\theta(a),a|x,z)\frac{\partial\tilde\theta(a)}{\partial t_1}da\\ &\quad -\int_{a^*}^{\overline{a}}\Big[t_2(x,z)-\tilde\theta(a)\mathrm{E}[P|2,x,z]\Big]\tilde f(\tilde\theta(a),a|x,z)\frac{\partial\tilde\theta(a)}{\partial t_1}da,\end{aligned}$$

which expresses the Lagrange multiplier in terms of observables.

Lemma 4 (Rationalization Lemma): *Let $\Psi(\cdot,\ldots,\cdot)$ be the distribution of $(J^*,D_1^*,\ldots,D_{J^*}^*,\chi,X,Z)$. Under Assumptions 1 and 2, $[F(\cdot,\cdot|\cdot,\cdot),H(\cdot|\cdot,\cdot)]\in\mathcal{F}_{XZ}\times\mathcal{H}_{XZ}$ rationalizes $\Psi(\cdot,\ldots,\cdot)$ if and only if the latter satisfies the following conditions*

(i) $\Psi_{D_1^*,\ldots,D_{J^*}^*|J^*,\chi,X,Z}(\cdot,\ldots,\cdot|\cdot,\ldots,\cdot)=\prod_{j=1}^{J^*}\Psi_{D_j^*|\chi,X,Z}(\cdot|\cdot,\ldots,\cdot)$, *where* $\Psi_{D_j^*|\chi,X,Z}(\cdot|\cdot,\cdot,\cdot)=\Psi_{D^*|\chi,X,Z}(\cdot|\cdot,\cdot,\cdot)\in\mathcal{H}^*_{\chi XZ}$,

(ii) For all $(x,z) \in \mathcal{S}_{XZ}$, $\psi_{D^*|\chi,X,Z}(\cdot|2,x,z)$ *and* $\psi_{D^*|\chi,X,Z}(\cdot|1,x,z)$ *are strictly positive on* $[dd_2(x,z),\overline{d}(x,z)]$ *and* $[dd_1(x,z),\overline{d}(x,z)]$*, respectively. Moreover, their ratio* $\pi(x,z)$ *is independent of* $d \in [dd_1(x,z),\overline{d}(x,z)]$ *with* $0 < \pi(x,z) < 1$,

(iii) For $c = 1,2$ *and all* $(x,z) \in \mathcal{S}_{XZ}$, $\psi_{J^*|\chi,X,Z}(\cdot|c,x,z) > 0$ *on* $I\!N$ *with a moment generating function defined on* $I\!R$ *such that the right-hand sides of (18) are the moment generating functions of absolutely continuous distributions with densities bounded away from zero on their supports* $[\underline{\tilde{\theta}}(1,x,z),\overline{\tilde{\theta}}(1,x,z)]$ *and* $[\underline{\tilde{\theta}}(2,x,z),\overline{\tilde{\theta}}(2,x,z)]$ *with union equal to* $[\underline{\tilde{\theta}}(1,x,z),\overline{\tilde{\theta}}(2,x,z)]$ *included in* $I\!R_{++}$.[13] *Moreover,* $\mathcal{S}_{a|\tilde{\theta}x} \equiv \{a : \exists z \in \mathcal{S}_{Z|\tilde{\theta}x}, a = \tilde{a}(\tilde{\theta},x,z)\}$ *is a compact interval in* $I\!R_{++}$ *independent of* $\tilde{\theta}$,

(iv) For every $(\tilde{\theta},x) \in \mathcal{S}_{\tilde{\theta}X}$

$$\left\{ \frac{\tilde{f}_{\tilde{\theta}|\chi,X,Z}[\tilde{\theta}|1,x,z]\psi_{\chi|X,Z}(1|x,z)}{\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x,z)}; z \in \mathcal{S}_{Z|\tilde{\theta}x} \right\} = [0,1],$$

(v) The coverage terms $t_1(\cdot,\cdot), t_2(\cdot,\cdot), dd_1(\cdot,\cdot), dd_2(\cdot,\cdot)$ *satisfy* $0 \le t_1(\cdot,\cdot) < t_2(\cdot,\cdot)$, $\overline{d}(\cdot,\cdot) \ge dd_1(\cdot,\cdot) > dd_2(\cdot,\cdot) \ge 0$, *and*

$$\begin{aligned}
&\int_{\underline{a}}^{a^*} \Big[t_1(x,z) - \tilde{\theta}(a)\mathrm{E}[P|1,x,z]\Big]\, \tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial dd_1}da + \mathrm{E}[J^*|1,x,z]\psi_{\chi,X,Z}(1,x,z) \\
&- \int_{a^*}^{\overline{a}} \Big[t_2(x,z) - \tilde{\theta}(a)\mathrm{E}[P|2,x,z]\Big]\, \tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial dd_1}da - \lambda(x,z)\underline{\tilde{\theta}}e^{\underline{a}dd_1(x,z)} = 0 \qquad (22) \\
&\int_{\underline{a}}^{a^*} \Big[t_1(x,z) - \tilde{\theta}(a)\mathrm{E}[P|1,x,z]\Big]\, \tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial t_2}da + \psi_{\chi|X,Z}(2|x,z)
\end{aligned}$$

[13] Alternatively, the conditions on the moment generating function of J^* given (χ,X,Z) in (iii) can be replaced by conditions on its characteristic function $\phi_{J^*|\chi,X,Z}(\cdot|c,x,z)$. Specifically, $\phi_{J^*|\chi,X,Z}(\cdot|c,x,z)$ is an entire characteristic function such that the right-hand sides of (29)-(30) are characteristic functions corresponding to absolutely continuous distributions with densities bounded away from zero on their supports $[\underline{\tilde{\theta}}(1,x,z),\overline{\tilde{\theta}}(1,x,z)]$ and $[\underline{\tilde{\theta}}(2,x,z),\overline{\tilde{\theta}}(2,x,z)]$ with union equal to $[\underline{\tilde{\theta}}(1,x,z),\overline{\tilde{\theta}}(2,x,z)]$ included in $I\!R_{++}$. Such conditions can be written equivalently in more testable forms. For instance, a function is a characteristic function if and only if it satisfies Bochner's Theorem 4.2.2, and it is entire if and only if it satisfies Theorem 7.2.1. A characteristic function corresponds to a distribution with bounded support in $I\!R_{++}$ if and only if it satisfies Theorem 7.2.3 with (7.2.3) strictly positive. All these theorems and equations are from Lukacs (1960). A well-known sufficient condition for a distribution to be absolutely continuous is that its characteristic function is absolutely integrable, while a necessary condition is that the characteristic function vanishes in the tails. See Billingsley (1995, pp.345-347).

$$-\int_{a^*}^{\overline{a}}\Big[t_2(x,z)-\tilde{\theta}(a)\mathrm{E}[P|2,x,z]\Big]\,\tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial t_2}da=0 \tag{23}$$

$$\int_{\underline{a}}^{a^*}\Big[t_1(x,z)-\tilde{\theta}(a)\mathrm{E}[P|1,x,z]\Big]\,\tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial dd_2}da+\mathrm{E}(J^*|\chi=2,x,z)\psi_{\chi,X,Z}(2|x,z)$$

$$-\int_{a^*}^{\overline{a}}\Big[t_2(x,z)-\tilde{\theta}(a)\mathrm{E}[P|2,x,z]\Big]\,\tilde{f}(\tilde{\theta}(a),a|x,z)\frac{\partial\tilde{\theta}(a)}{\partial dd_2}da=0 \tag{24}$$

$$t_1(x,z)=\frac{\tilde{\underline{\theta}}}{\underline{a}}\left[\int_{dd_1(x,z)}^{\overline{d}(x,z)}\left(e^{\underline{a}D}-e^{\underline{a}dd_1(x,z)}\right)\psi_{D^*|\chi,X,Z}(D|1,x,z)dD\right]. \tag{25}$$

Condition (i) says that reported damages are independent and identically distributed given the coverage choice and individual/car characteristics. In addition, reported damages are independent of the reported number of accidents given those variables. This is a consequence of Assumption 1-(i,ii) on damages and number of accidents. Condition (ii) requires that the densities of reported damages given coverage choice and individual/car characteristics are strictly positive on their supports. More importantly, the ratio of these densities needs to be independent of the level of reported damage above the level of highest deductible following (15). This property is also a consequence of Assumption 1-(i,ii), i.e. independently and identically distributed damages and independence of damages from coverage choice. Condition (iii) states that the support of the distribution of reported accidents given coverage choice and individual/car characteristics is the set of integers. The remaining part of (iii) follows from the compact support of the joint distribution of risk and risk aversion and its nonvanishing corresponding density by Definition 1. In view of (21), condition (iv) says that the probability for choosing coverage 1 by an insuree characterized by (θ, x, z) takes all values in $[0,1]$ as the car characteristics vary. This follows from the full support assumption in Assumption 2. Condition (v) provides the relationship between the distribution of observables and the coverage terms. In particular, it requires that the premium and deductible for the two coverages must satisfy the optimality conditions, i.e. the first-order conditions (8)-(12).

The rationalization lemma is important for several reasons. First, the insurance model with multidimensional private information does impose some restrictions on observables. In view of bunching in our model due to multidimensional screening and a finite number of coverages, one could have expected otherwise a priori. For instance, in auction models, the main restriction arises from the monotonicity of the equilibrium bidding strategy, which is not present here because of the finite number of contracts. Second, Lemma 4 characterizes all the restrictions on the distribution of observables. It can be used to test the validity of the model. Violation of a single restriction by the data would reject the model. Relying on some recent developments in the econometrics literature, we can discuss some testing procedures for each condition. For instance, (i) can be implemented using conditional independence tests. See Su and White (2008). The independence of $\pi(x,z)$ from damage can be tested by noting that the ratio of the densities is also equal to the ratio $\Psi_{D^*|\chi,X,Z}(dd_1(x,z)|2,x,z)/\Psi_{D^*|\chi,X,Z}(dd_1(x,z)|1,x,z)$. This equality can then be used to derive a Cramer-Van Mises type test relying on nonparametric estimates of the densities. See Brown and Wegkamp (2002). Regarding (iii), as noted in footnote 12, we can equivalently derive the restrictions that must satisfy the corresponding characteristic function because the moment generating function at u is equal to the characteristic function at $-iu$. See Lukacs (1960, p. 135). Such restrictions would be then more convenient to derive appropriate tests of such conditions.

Third, (v) provides restrictions on the coverage terms suggesting that an optimality test could be performed. This contrasts with the previous structural literature in which it is generally assumed that the observations are the outcomes of some equilibrium. In particular, in auctions, observed bids result from the Bayesian Nash equilibrium of the auction game and identification relies on such optimal behavior. This represents a strong assumption that might be questionable from an empirical point of view. In contrast, when the number of contracts is finite, optimality of the coverage terms given by (8)–(12) is not used in identifying the model structure. As noted earlier, we exploit only the optimal partitioning of insurees among the two contracts taking the observed premiums and deductibles as given leading to the restrictions (i)–(iv). Thus, (22)–(25) can be used to test the optimality of the observed coverages (T_1, DD_1, T_2, DD_2).[14]

[14] Though we use a different identification strategy for a continuum of contracts relying on the monotonicity of the coverage terms in the certainty equivalence, we can envision a similar identification strategy as in Section 4.2. This suggests that we could test the validity of the contract terms from

6 Identification Strategies for Case 4

From Section 4.2, any assumption that identifies $H_2(X, Z)$ will identify the structure $[F(\cdot, \cdot|X, Z), H(\cdot|X, Z)]$ on $\Theta(X, Z) \times \mathcal{A}(X, Z)$ and $[dd_2(X, Z), \overline{d}(X, Z)]$, respectively. In this section, we investigate some identifying assumptions/conditions for $H_2(X, Z)$. Another possibility is to derive some bounds for $H_2(X, Z)$.

Parameterization of $H(\cdot|X, Z)$

A simple stategy to identify $H_2(X, Z)$ is to parameterize the damage distribution $H(\cdot|X, Z)$ as $H(\cdot|X, Z; \gamma)$ on $[0, \overline{d}(X, Z)]$ with $\gamma \in \Gamma \subset I\!R^q$. Observations on reported damage D^* will typically identify γ and hence $H(\cdot|X, Z)$ on $[0, \overline{d}(X, Z)]$. In particular, $H_2(X, Z) = H(dd_2(X, Z)|X, Z; \gamma)$ will be identified. So far, we have tried to minimize parametric assumptions. From an estimation point of view, one could estimate nonparametrically the truncated damage conditional density and use its shape to choose the parameterization of $H(\cdot|X, Z)$. This exercise would require some reasonable assumptions on the damage distribution such as continuity on its support and no mass point below $dd_2(X, Z)$.

Additional Data Sources

A second strategy is to consider additional data sources providing for instance the average number of accidents (reported and unreported) for every $(x, z) \in \mathcal{S}_{XZ}$, i.e. $\mu(x, z) = \mathrm{E}[\theta|X = x, Z = z]$. Let the average number of reported accidents for every (x, z) be $\mu_c^*(x, z) = \mathrm{E}(\theta|\chi = c, X = x, Z = z)(1 - H_c(X, Z))$ for $c = 1, 2$. We have

$$\begin{aligned}
\mu(x, z) &= \nu_1(x, z)\mathrm{E}[\theta|\chi = 1, X = x, Z = z] + \nu_2(x, z)\mathrm{E}[\theta|\chi = 2, X = x, Z = z] \\
&= \frac{1}{1 - H_2(x, z)}\left(\nu_1(x, z)\frac{\mu_1^*(x, z)}{\pi(x, z)} + \nu_2(x, z)\mu_2^*(x, z)\right)
\end{aligned}$$

leading to the identification of $H_2(x, z)$ given that $\nu_c(x, z)$, $\mu_c^*(x, z), c = 1, 2$ and $\pi(x, z)$ are identified from the data as shown in Section 4.2. Alternatively, an auxiliary information could be $\mathrm{E}(\theta|\chi = c, X = x, Z = z)$ for (say) $c = 2$ and every (x, z). From the knowledge of $\mu_2^*(x, z)$, it is straightforward to identify $H_2(x, z)$.

---

the first-order conditions (5)-(6) even when a continuum of coverages is offered albeit an assumption similar to Assumption 2 has to be made.

Next, we consider that an auxiliary information is $\mathrm{E}[\theta|X = x_0, Z = z_0]$ for some (x_0, z_0). Using the argument in the previous paragraph shows that $H_2(x_0, z_0)$ is identified. This information combined with a support assumption such as $\overline{\theta}(x, z) = \overline{\theta}$ for every (x, z) identifies $H_2(x, z)$. Specifically, note that we have $\overline{\tilde{\theta}}(x, z) = (1 - H_2(x, z))\overline{\theta}(x, z)$, where $\overline{\tilde{\theta}}(x, z)$ is the upper boundary of the support of $f_{\tilde{\theta}|X,Z}(\cdot|X = x, Z = z)$, which is identified as shown in Section 4.2. Applying this equation at (x_0, z_0) identifies $\overline{\theta}$ by $\overline{\tilde{\theta}}(x_0, z_0)/(1 - H_2(x_0, z_0))$. Applying again this equation at different values (x, z) identifies $H_2(x, z)$. A similar argument applies if $\underline{\theta}(x, z) = \underline{\theta}$.

It remains to investigate whether additional information on damages (reported and unreported) helps in identifying $H_2(x, z)$. We have

$$\begin{aligned}\mathrm{E}(D|X = x, Z = z) \;\; = \;\; & \mathrm{E}[D|D \le dd_2(x, z)|X = x, z = z]H_2(x, z) \\ & +\mathrm{E}[D|D \ge dd_2(x, z)|X = x, z = z](1 - H_2(x, z)),\end{aligned}$$

where $\mathrm{E}[D|D \ge dd_2(x, z), X = x, z = z] = \int_{dd_2(x,z)}^{\overline{d}(x,z)} Dh_2^*(D|X = x, Z = z)dD$ is identified from the data. Thus, for every (x, z) it is straightforward to see that identification of $H_2(x, z)$ requires to know both $\mathrm{E}[D|D \le dd_2(x, z), X = x, Z = z]$ and $\mathrm{E}(D|X = x, Z = z)$. In particular, the knowledge of the latter is not sufficient in constrast to the previous case in which additional data on the average number of accidents only was sufficient for identification. As above, if one knows $\mathrm{E}[D|D \le dd_2(x_0, z_0), X = x_0, Z = z_0]$ and $\mathrm{E}(D|X = x_0, Z = z_0)$ for some (x_0, z_0) and if either $\overline{\theta}(x, z)$ or $\underline{\theta}(x, z)$ is independent of (x, z), then $H_2(x, z)$ is identified for every (x, z).

Set Identification

A third strategy is to derive some bounds on $H_2(X, Z)$, which will provide some bounds on the structure $[F(\cdot, \cdot|X, Z), H(\cdot|X, Z)$. This approach also known as set identification has been made popular by Manski and Tamer (2002) and Chernozhukov, Hong and Tamer (2007). See also Haile and Tamer (2003) and Kovchegov and Yildiz (2009) for nonparametric bounds. Our bounds are in the spirit of the latter as they are nonparametric. Let $[F^0(\cdot, \cdot|X, Z), H^0(\cdot|X, Z)]$ be the true structure. Given an arbitrary pair of values (x, z), Proposition 4 implies that it is sufficient to determine the identified set for

$H_2^0(x,z)$, i.e. the set of values $H_2(x,z)$ that are observationally equivalent to $H_2^0(x,z)$.[15] The proof of Lemma 3 shows that any value $H_2(x,z) = 1-(1/\kappa)[1-H_2^0(x,z)]$ for $\kappa > \sup_{(\tilde{x},\tilde{z})}[1-H_2^0(\tilde{x},\tilde{z})]$ is observationally equivalent to $H_2^0(x,z)$. Thus, the identified set for $H_2^0(x,z)$ contains the interval

$$\left(1-\frac{1-H_2^0(x,z)}{\sup_{(\tilde{x},\tilde{z})}[1-H_2^0(\tilde{x},\tilde{z})]},1\right). \tag{26}$$

For the values (x,z) for which $1-H_2^0(x,z)$ is close to the supremum, the left boundary of the above interval approaches zero. Hence, for those values, the identified set is close to $(0,1)$, which is not informative.

Some empirical evidence in Cohen and Einav (2007) may help us to motivate an additional assumption that renders these bounds tighter. In particular, their estimated damage density strictly decreases when the damage approaches the deductible from above suggesting that the density below the deductible is not greater than its value at the deductible. We then make the following assumption.

Assumption 3: *The conditional damage distribution $H(\cdot|X,Z)$ satisfies*

$$h(D|x,z) \le h[dd_2(x,z)|x,z],$$

for every $D \le dd_2(x,z)$ and $(x,z) \in \mathcal{S}_{XZ}$.

We use this assumption to construct more informative bounds. Specifically, integrating both sides from 0 to $dd_2(x,z)$ we obtain $0 \le H_2(x,z) \le dd_2(x,z)h(dd_2(x,z)|x,z)$. Dividing both sides by $1-H_2(x,z)$ and using the definition of the conditional density $h_2^*(\cdot|x,z)$, we obtain

$$0 \le \frac{H_2(x,z)}{1-H_2(x,z)} \le dd_2(x,z)h_2^*(dd_2(x,z)|x,z).$$

Solving for $H_2(x,z)$ gives the bounds

$$0 \le H_2(x,z) \le \frac{dd_2(x,z)h_2^*(dd_2(x,z)|x,z)}{1+dd_2(x,z)h_2^*(dd_2(x,z)|x,z)} \equiv \overline{B}(x,z). \tag{27}$$

[15] To be precise, this is the set of values $H_2(x,z)$ corresponding to structures $[F(\cdot,\cdot|X,Z),H(\cdot|X,Z)]$ that are observationally equivalent to $[F^0(\cdot,\cdot|X,Z),H^0(\cdot|X,Z)]$.

In particular, the upper bound for $H_2(x,z)$ is strictly less than 1. Moreover, a useful feature of the bounds (27) is that they are expressed as functions of observables.[16]

It remains to derive some bounds on the structure $[F(\cdot,\cdot|X,Z), H(\cdot|X,Z)]$. From (19) we obtain the following lower and upper bounds for $H(\cdot|x,z)$

$$[H_2^*(\cdot|x,z), H_2^*(\cdot|x,z) + \overline{B}(x,z)(1 - H_2^*(\cdot|x,z))]$$

for every $\cdot \geq dd_2(x,z)$ and $(x,z) \in \mathcal{S}_{XZ}$. Thus, the smaller is $\overline{B}(x,z)$, i.e. the smaller is $dd_2(x,z)h_2^*(dd_2(x,z)|x,z)$, the narrower is the above interval.

Regarding the derivation of bounds on $F(\cdot,\cdot|X,Z)$, we follow the identification argument of Section 4.2. We first derive bounds for the marginal c.d.f of θ given (X,Z). Recall that the c.d.f. of $\tilde{\theta} = (1 - H_2(x,z))\theta$ is identified from its moment generating function and the observed number of reported accidents. In particular, we have $F_{\theta|X,Z}(\cdot|x,z) = \tilde{F}_{\tilde{\theta}|X,Z}[(1 - H_2(x,z)) \cdot |x,z]$ showing that

$$\tilde{F}_{\tilde{\theta}|X,Z}[(1 - \overline{B}(x,z)) \cdot |x,z] \leq F_{\theta|X,Z}(\cdot|x,z) \leq \tilde{F}_{\tilde{\theta}|X,Z}(\cdot|x,z),$$

leading to a first-order stochastic dominance among these three c.d.f.s. Section 4.2 does not provide, however, an explicit expression for the c.d.f. $\tilde{F}_{\tilde{\theta}|X,Z}(\cdot|x,z)$ as its identification was established through its moment generating function (19).

To obtain such an explicit form, we consider its density $\tilde{f}_{\tilde{\theta}|X,Z}(\cdot|x,z)$ and determines its characteristic function $\phi_{\tilde{\theta}|X,Z}(\cdot|x,z)$ from available data. We first remark that the distribution of $\tilde{\theta}$ given (χ, X, Z) has compact support. Thus, it has an entire characteristic function $\phi_{\tilde{\theta}|\chi,X,Z}(\cdot|c,x,z)$, i.e. a characteristic function that has a (unique) differentiable extension on the whole set of complex numbers $\phi_{\tilde{\theta}|\chi,X,Z}(\cdot|c,x,z) = \mathrm{E}[e^{i\zeta\tilde{\theta}}|\chi = c, X =$

[16] To show that the bounds (27) are sharp requires to obtain the set of observationally equivalent values $H_2(x,z)$, and in particular, the sharp lower bound of this set. The previous discussion shows that the latter is between 0 and the lower bound of the interval (26). Similarly, exploiting the relationship $1 - H_2(x,z) = [1 - H_1(x,z)]/\pi(x,z)$ we obtain bounds for $H_1(x,z)$, namely

$$1 - \pi(x,z) \leq H_1(x,z) \leq 1 - \frac{\pi(x,z)}{1 + dd_2(x,z)h_2^*(dd_2(x,z)|x,z)}.$$

The lower and upper bounds for $H_1(x,z)$ are strictly larger than zero and smaller than one, respectively.

$x, Z = z]$ for $\zeta \in \mathbb{C}$. See Lukacs (1960, p. 139). Following the derivation leading to (16) with t replaced by $i\zeta$ and noting that the characteristic function of a Binomial $\mathcal{B}(n,p)$ and a Poisson $\mathcal{P}(\lambda)$ random variables are entire with extensions equal to $(1-p+pe^{i\zeta})^n$ and $e^{\lambda(e^{i\zeta}-1)}$, where $\zeta \in \mathbb{C}$, we obtain

$$\begin{aligned} \phi_{J^*|\chi,X,Z}(\zeta|c,x,z) &= \phi_{\theta|\chi,X,Z}\left[(1-H_c(x,z))\frac{e^{i\zeta}-1}{i}|c,x,z\right] \\ &= \phi_{\tilde{\theta}|\chi,X,Z}\left[\frac{1-H_c(x,z)}{1-H_2(x,z)}\frac{e^{i\zeta}-1}{i}|c,x,z\right], \end{aligned} \tag{28}$$

where the second equality follows from $\tilde{\theta} = (1-H_2(X,Z))\theta$. Let $\tilde{\zeta} = [(1-H_c(x,z))(e^{i\zeta}-1)]/[i(1-H_2(x,z))]$, for all $\zeta \in \mathbb{C}$ and $\zeta = u + i\log(\cos u)$ for $u \in (-\pi/2, \pi/2)$. Then, $\tilde{\zeta} = \tan u$ when $c = 2$ and $\tilde{\zeta} = \pi(x,z)\tan u$ when $c = 1$. Moreover, the range of $\tilde{\zeta}$ is $\mathbb{R}$. Therefore, letting $t = \arctan u$ when $c = 2$ and $t = \pi(x,z)\arctan u$ when $c = 1$ in (28), and using $\cos(\arctan t) = 1/\sqrt{1+t^2}$ give the characteristic functions

$$\begin{aligned} \phi_{\tilde{\theta}|\chi,X,Z}(t|2,x,z) &= \phi_{J^*|\chi,X,Z}\left[\arctan t - \frac{i}{2}\log(1+t^2)|2,x,z\right] \\ \phi_{\tilde{\theta}|\chi,X,Z}(t|1,x,z) &= \phi_{J^*|\chi,X,Z}\left[\arctan\left(\frac{t}{\pi(x,z)}\right) - \frac{i}{2}\log\left(1+\frac{t^2}{\pi^2(x,z)}\right)|1,x,z\right], \end{aligned}$$

for all $t \in \mathbb{R}$. Since $\phi_{\tilde{\theta}|X,Z}(t|x,z) = \nu_1(x,z)\phi_{\tilde{\theta}|\chi,X,Z}(t|1,x,z) + \nu_2(x,z)\phi_{\tilde{\theta}|\chi,X,Z}(t|2,x,z)$, one obtains the density $\tilde{f}_{\tilde{\theta}|X,Z}(\cdot|\cdot,\cdot)$ by the inverse Fourier transform

$$\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x,z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-it\tilde{\theta}}\phi_{\tilde{\theta}|X,Z}(t|x,z)dt.$$

if $\phi_{\tilde{\theta}|X,Z}(\cdot|x,z)$ is absolutely integrable, in which case $\tilde{f}_{\cdot|X,Z}(\tilde{\theta}|x,z)$ is continuous. See e.g. Billingsley (1995, pp.347-348). Lastly, to determine the identified set for $F_{a|\theta,X}(\cdot|\cdot,\cdot)$, one can use the bounds for $H_2(X,Z)$ and follow the identifying argument of Section 4.2.

7 Conclusion

Our paper addresses the problem of identification of insurance models with multidimensional screening, where insurees have private information on their risk and risk aversion, each taking a continuum of values. We define risk as the expected number of accidents.

We make a special effort to incorporate in our model some important features in insurance such as a random damage and the possibility of several accidents. The model also considers the possibility that the contracts offered to each insuree is either a continuum or finite in number. Moreover, we also allow for data restrictions on the number of accidents and their corresponding damages as insurees are expected to report an accident only when the damage is above the deductible. Bunching arises necessarily at the equilibrium because of mutlidimensional private information. The bunching problem is accentuated when there is a finite number of coverages. Consequently, insurees with different private information can choose the same coverage, which complicates the problem of identification from coverage choices. Despite this, we show that we identify the joint distribution of risk and risk aversion. When a continuum of contracts is offered, identification is achieved without any additional assumption by exploiting the number of (reported) accidents. When a finite number of contracts is offered, this information also plays a crucial role though additional identifying assumptions need to be made. When only reported accidents and damages are available to the analyst, we provide several identifying strategies including set identification. Lastly, we characterize all the restrictions imposed by the model on observables. Such restrictions can be used to test the model validity. An interesting feature of these restrictions is that optimality of the offered coverages can be tested separately as identification of the model does not rely on this property.

Our results can readily be used to analyze insurance data when a limited number of coverages is offered to each insuree and accidents are reported only when the damage is above the deductible. The estimation method can follow the identification steps. Section 6 provides the material needed to estimate nonparametrically the model. For instance, one could use nonparametric density estimators to obtain estimates of the bounds for the probability of damage below the deductible. Inverse Fourier transform of the empirical characteristic function can then be used to estimate bounds on the marginal distribution of risk. Our nonparametric approach leaves much flexibility on the degree of dependence between risk and risk aversion. Our companion paper develops this estimation procedure with an application to automobile insurance data. See Aryal, Perrigne and Vuong (in progress).

Our model allows the individual and car characteristics to be dependent. This can be exploited further by endogeneizing the car choice given the individual's characteristics, risk and risk aversion. This would lead to a model explaining the car choice, the coverage choice, the number of accidents and the damages. Our results still hold in this case. In particular, the identification of the joint distribution of risk and risk aversion can help in estimating and/or simulating this model. See Aryal and Perrigne (in progress). More generally, our results can be used to analyze a large range of insurance data such as crop, health or home insurance as long as the data provide a repeated outcome such as several claims made by insurees. Our identification method can also be used to study identification of models with adverse selection when a finite number of contracts is offered. One can think of nonlinear pricing where a finite number of price/quantity options is offered to consumers or firms who purchase items several times over the pricing period.

Appendix

Proof of Lemma 3: In view of Proposition 4, $H_2(X,Z)$ is identified if and only if the structure $[F(\cdot,\cdot|X,Z),H(\cdot|X,Z)]$ is. Thus, it suffices to show that the latter is not identified. Let $[F(\cdot,\cdot|X,Z),H(\cdot|X,Z)]$ be a structure satisfying Definitions 1 and 2 as well as Assumptions 1 and 2. We construct a second structure $[\tilde{F}(\cdot,\cdot|X,Z),\tilde{H}(\cdot|X,\ Z)]$ as follows. Let $\tilde{\theta}=\kappa\theta$ with $\kappa>\sup_{(x,z)\in\mathcal{S}_{XZ}}[1-H_2(x,z)]\geq 0$, while $\tilde{a}=a$ so that $\tilde{f}(\cdot,\cdot|X,Z)=(1/\kappa)f(\cdot/\kappa,\cdot|X,Z)$. Let $\tilde{h}(\cdot|X,Z)$ be a strictly positive conditional density on its support $[0,\overline{d}(X,Z)]$ with $\tilde{h}(D|X,Z)=(1/\kappa)h(D|X,Z)$ for $D\geq dd_2(X,Z)$. Because $0<\int_{dd_2(x,z)}^{\overline{d}(x,z)}\tilde{h}(D|x,z)dD<1$, it follows that $\kappa>1-H_2(x,z)$ for all $(x,z)\in\mathcal{S}_{XZ}$ as required above. The second structure $[\tilde{F}(\cdot,\cdot|X,Z),\tilde{H}(\cdot|X,Z)]$ satisfies Definitions 1 and 2 as well as Assumptions 1 and 2 as $\tilde{\theta}(a,X,Z)=\kappa\theta(a,X,Z)$ as shown below.

We now show that these two structures are observationally equivalent, i.e. they lead to the same distribution for the observables $(J^*,D_1^*,\ldots,D_{J^*}^*,\chi,t_1,dd_1,t_2,dd_2)$ given (X,Z), where J^* and D^* refer to the number of reported accidents and their corresponding damages, respectively, while χ indicates which coverage is chosen by the insuree. First, we note that the coverage terms are deterministic functions of (X,Z) solving the FOC (8)–(12). Thus, the optimal frontier for the second structure must be

$$\begin{aligned}\tilde{\theta}(a,X,Z) &= \frac{t_2(X,Z)-t_1(X,Z)}{\int_{dd_2(X,Z)}^{dd_1(X,Z)}e^{aD}(1-\tilde{H}(D|X,Z))dD}=\frac{t_2(X,Z)-t_1(X,Z)}{\int_{dd_2(X,Z)}^{dd_1(X,Z)}e^{aD}\frac{1}{\kappa}(1-H(D|X,Z))dD}\\ &= \kappa\theta(a,X,Z),\end{aligned}$$

thereby showing that the highest risk aversion in $\tilde{\mathcal{A}}_1$ is $\tilde{a}^*(X,Z)=a^*(X,Z)$.

Regarding the distribution $\tilde{\chi}$ given (X,Z), we note that $\tilde{\chi}=\chi$. The latter follows from $\tilde{\chi}=1$ if and only if $(\tilde{\theta},a)\in\tilde{\mathcal{A}}_1(X,Z)$, i.e. $\tilde{\theta}\leq\tilde{\theta}(a,X,Z)$ and $\underline{a}(X,Z)\leq a\leq\tilde{a}^*(X,Z)$. Since $\tilde{\theta}=\kappa\theta$, $\tilde{\theta}(a,X,Z)=\kappa\theta(a,X,Z)$ and $\tilde{a}^*(X,Z)=a^*(X,Z)$, we have $\tilde{\chi}=1$ if and only if $\chi=1$. Thus, the distribution of $\tilde{\chi}$ given (X,Z) is the same as that of χ given (X,Z), i.e. $\tilde{\nu}_c(X,Z)=\nu_c(X,Z)$ for $c=1,2$. Regarding the distribution of $\tilde{J}^*$ given $(\tilde{\chi},X,Z)=(\chi,X,Z)$, from (16) its moment generating function is

$$\begin{aligned}M_{\tilde{\theta}|\chi,X,Z}[(1-\tilde{H}_\chi(X,Z))(e^t-1)|c,x,z] &= M_{\theta|\chi,X,Z}[(1-H_\chi(X,Z))(e^t-1)|c,x,z]\\ &= M_{J^*|\chi,X,Z}[t|c,x,z]\end{aligned}$$

using $1-\tilde{H}_c(X,Z)=(1-H_c(X,Z))/\kappa$ and $M_{\tilde{\theta}|\chi,X,Z}(u|c,x,z)=M_{\theta|\chi,X,Z}(\kappa u|c,x,z)$. Hence, the distribution of $\tilde{J}^*$ given (χ,X,Z) is the same as that of J^* given (χ,X,Z). Regarding the distribution of reported damage $\tilde{D}^*$ given $(\tilde{J}^*,\chi,X,Z)$ is

$$\tilde{H}^*_\chi(\cdot|X,Z)=\frac{\tilde{H}(\cdot|X,Z)-\tilde{H}_\chi(X,Z)}{1-\tilde{H}_\chi(X,Z)}=\frac{H(\cdot|X,Z)-H_\chi(X,Z)}{1-H_\chi(X,Z)}=H^*_\chi(\cdot|X,Z)$$

using $1-\tilde{H}_\chi(\cdot|X,Z)=(1-H_\chi(\cdot|X,Z))/\kappa$.

Lastly, it remains to show that $(t_1(X,Z),dd_1(X,Z),t_2(X,Z),dd_2(X,Z))$ satisfies the FOC (8)–(12) associated with the second structure. Using $\tilde{\theta}(a,X,Z)=\kappa\theta(a,X,Z)$, $\tilde{f}(\tilde{\theta}(a,X,Z),a|X,Z)=f(\tilde{\theta}(a,X,Z)/\kappa,a|X,Z)/\kappa=f(\theta(a,X,Z),a|X,Z)/\kappa$, $1-\tilde{H}(D|X,Z)=(1-H(D|X,Z))/\kappa$, $\tilde{\nu}_c=\nu_c$ and $\mathrm{E}[\tilde{\theta}|\tilde{\mathcal{A}}_c]=\kappa\mathrm{E}[\theta|\mathcal{A}_c]$, it can be easily verified that $(t_1(X,Z),dd_1(X,Z),t_2(X,Z),dd_2(X,Z))$ satisfies (8)–(12) with $\tilde{\lambda}=\lambda$ as soon as (8)–(12) hold for the original structure. Hence, the two structures lead to the same distributions for the observables as desired.□

Proof of Lemma 4: We first prove neccessity. Let $[F(\cdot,\cdot|\cdot,\cdot),H(\cdot|\cdot,\cdot)]\in\mathcal{F}_{XZ}\times\mathcal{H}_{XZ}$ be a structure that rationalizes $\Psi(\cdot,\ldots,\cdot)$ under Assumptions 1 and 2. To prove (i) we follow Guerre, Perrigne and Vuong (2000) proof of Theorem 4 (Conditions C1-C2). From Assumption 1-(i,ii), we have $(D_1,\ldots,D_J)$ i.i.d as $H(\cdot|X,Z)$ conditional upon (J,θ,a,X,Z). Thus, J^* follows a $\mathcal{B}[J,1-H_\chi(X,Z)]$ given (J,θ,a,X,Z) since an accident is reported if and only if the damage is above the deductible. For any $(d_1,\ldots,d_j)\in\mathbb{R}^j_+$ we have

$$\begin{aligned}
&\Pr[D^*_1\le d_1,\ldots,D^*_j\le d_j,J^*=j|J,\theta,a,X,Z]\\
&=\sum_{1\le r_1\neq\ldots\neq r_j\le J}\Pr[dd_\chi(X,Z)\le D_{r_1}\le d_1,\ldots,dd_\chi(X,Z)\le D_{r_j}\le d_j,D_r<dd_\chi(X,Z),r\notin\{r_1,\ldots,r_j\}|J,\theta,a,X,Z]\\
&=\frac{J!}{j!(J-j)!}\Pr[dd_\chi(X,Z)\le D_1\le d_1,\ldots,dd_\chi(X,Z)\le D_j\le d_j,D_r<dd_\chi(X,Z),r=j+1,\ldots,J|J,\theta,a,X,Z]\\
&=\frac{J!}{j!(J-j)!}\left(\prod_{r=1}^{j}[H(d_r|X,Z)-H_\chi(X,Z)]\right)[H_\chi(X,Z)]^{J-j}
\end{aligned}$$

because $(D_1,\ldots,D_J)$ are i.i.d. as $H(\cdot|X,Z)$ given (J,θ,a,X,Z). Since J^* is $\mathcal{B}[J,1-H_\chi(X,Z)]$ given (J,θ,a,X,Z) we obtain

$$\Pr[D^*_1\le d_1,\ldots,D^*_j\le d_j|J^*=j,J,\theta,a,X,Z]=\prod_{r=1}^{j}\frac{H(d_r|X,Z)-H_\chi(X,Z)}{1-H_\chi(X,Z)}$$

showing that $(D^*_1,\ldots,D^*_j)$ are i.i.d as $H^*_\chi(X,Z)\in\mathcal{H}^*_{\chi XZ}$ given $(J^*=j,J,\theta,a,X,Z)$ and hence given $(J^*=j,\chi,X,Z)$. Thus, (i) holds.

To prove (ii), we note that $\Psi_{D^*|\chi,X,Z}(\cdot|\cdot,\cdot,\cdot) = H^*_\chi(\cdot,\cdot) \in \mathcal{H}^*_{\chi XZ}$ thereby establishing the first part of (ii). Moreover, $\psi_{D^*|\chi,X,Z}(d|2,x,z)/\psi_{D^*|\chi,X,Z}(d|1,x,z) = (1-H_1(x,z))/(1-H_2(x,z)) \equiv \pi(x,z)$, which is independent of $d \in [dd_1(x,z), \overline{d}(x,z)]$ and in $(0,1)$.

To prove (iii), we note that

$$\Pr[J^* = j^*|\theta,a,X,Z] = \sum_{j=j^*}^{\infty} \Pr[J^* = j^*|J=j,\theta,a,X,Z]\Pr[J=j|\theta,a,X,Z].$$

Thus, J^* given (θ,a,X,Z) is a mixture of a $\mathcal{B}[J, 1-H_\chi(X,Z)]$ with a mixing $\mathcal{P}(\theta)$ distribution by Assumption 1-(iii). That is, $\Psi_{J^*|\theta,a,X,Z}(\cdot|\theta,a,x,z)$ is a $\mathcal{P}[(1-H_\chi(x,z))\theta]$ distribution. Hence, $\psi_{J^*|\chi,X,Z}(\cdot|c,x,z) = \int_{\mathcal{A}_c} \Psi_{J^*|\theta,a,X,Z}(\cdot|\theta,a,x,z)dF(\theta,a|x,z)$ thereby establishing $\psi_{J^*|\chi,X,Z}(\cdot|c,x,z) > 0$ on $I\!N$ as $F(\cdot,\cdot|\cdot,\cdot) \in \mathcal{F}_{XZ}$. The moment generating function of J^* given (χ,X,Z) exists on $I\!R$ in view of (16) since the distribution of θ given (χ,X,Z) has a bounded support. The right-hand sides of (18) must be the moment generating functions of absolutely continuous distributions with densities bounded away from zero on their supports $[\underline{\tilde{\theta}}(1,x,z), \overline{\tilde{\theta}}(1,x,z)]$ and $[\underline{\tilde{\theta}}(2,x,z), \overline{\tilde{\theta}}(2,x,z)]$ with union equal to $[\underline{\tilde{\theta}}(1,x,z), \overline{\tilde{\theta}}(2,x,z)]$ included in $I\!R_{++}$ because they are the moment generating functions of $\tilde{\theta} = (1-H_2(X,Z))\theta$ given (c,x,z) which have such properties.

Regarding (iv), for every $(\theta,a,x) \in \mathcal{S}_{\theta aX}$, we have

$$\begin{aligned} F_{a|\theta,X}(a|\theta,x) &= F_{a|\theta,X,Z}[a(\theta,x,z)|\theta,x,z] = \frac{f_{\theta|\chi,X,Z}(\theta|1,x,z)\psi_{\chi|x,z}(1|x,z)}{f_{\theta|X,Z}(\theta|x,z)}, \\ &= \frac{\tilde{f}_{\tilde{\theta}|\chi,X,Z}(\tilde{\theta}|1,x,z)\psi_{\chi|x,z}(1|x,z)}{\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x,z)}, \end{aligned}$$

for some $z \in \mathcal{S}_{Z|\theta x}$, where the first equality follows from Assumption 2, the second equality from Bayes rule, and the third equality from $\tilde{\theta} = (1-H_2(X,Z))\theta$. Because a can be chosen arbitrarily, it follows that the right-hand side takes all values in $[0,1]$. Regarding (v), let $\tilde{\theta} = (1-H_2(X,Z))\theta$. The proof then follows the last paragraph of the proof of Lemma 3 with $\kappa = 1 - H_2(X,Z)$.

We now turn to sufficiency. Let the distribution $\Psi(\cdot,\ldots,\cdot)$ of $(J^*, D^*_1, \ldots, D^*_{J^*}, \chi, X, Z,)$ and the contract terms $[t_1(\cdot,\cdot), dd_1(\cdot,\cdot), t_2(\cdot,\cdot), dd_2(\cdot,\cdot)]$ satisfy (i)–(v). We need to exhibit a structure $[F(\cdot,\cdot|\cdot,\cdot), H(\cdot|\cdot,\cdot)] \in \mathcal{F}_{XZ} \times \mathcal{H}_{XZ}$ satisfying Assumptions 1–2 that rationalizes $\Psi(\cdot,\ldots,\cdot)$ of $(J^*, D^*_1, \ldots, D^*_{J^*}, \chi, X, Z,)$ and $[t_1(\cdot,\cdot), dd_1(\cdot,\cdot), t_2(\cdot,\cdot), dd_2(\cdot,\cdot)]$. Let the distribution of $(J, D_1, \ldots, D_J, \theta, a, X, Z)$ satisfy Assumptions 1 and 2-(i).

In view of the identification argument of Section 4.2, we define $H(\cdot|\cdot,\cdot)$ as follows: For a constant $\kappa \in (0,1)$, let $H(D|X,Z) = \kappa\psi_{D^*|\chi,X,Z}(D|2,X,Z)+(1-\kappa)$ when $D \geq dd_2(X,Z)$. Note that

$H(\cdot|X,Z)$ has a strictly positive density on $[dd_2(X,Z), \overline{d}(X,Z)]$ because $\Psi_{D^*|\chi,X,Z}(\cdot|2,X,Z) \in \mathcal{H}^*_{2XZ}$. For $D \in [0, dd_2(X,Z)]$ let $H(\cdot|X,Z)$ be arbitrary as long as it has a strictly positive density on $[0, dd_2(X,Z)]$. Thus, $H(\cdot|\cdot,\cdot) \in \mathcal{H}_{X,Z}$. Note that $\kappa = 1 - H(dd_2(X,Z)|X,Z) \equiv 1 - H_2(X,Z)$ so that $H^*_2(\cdot|X,Z) \equiv [H(\cdot|X,Z) - H_2(X,Z)]/[1 - H_2(X,Z)] = \Psi_{D^*|\chi,X,Z}(\cdot|2,X,Z)$ after straightforward algebra. Moreover, $\psi_{D^*|\chi,X,Z}(D|2,X,Z) = \pi(X,Z)\ \psi_{D^*|\chi,X,Z}(D|1,X,Z)$ for $D \geq dd_1(X,Z)$ by (ii) implying $\pi(X,Z) = 1 - \Psi_{D^*|\chi,X,Z}[dd_1(X,Z)|2,X,Z]$ by integration and $H^*_1(\cdot|X,Z) \equiv [H(\cdot|X,Z) - H_1(X,Z)]/[1 - H_1(X,Z)] = \Psi_{D^*|\chi XZ}(\cdot|1,X,Z)$ after some algebra. Thus, $\Psi_{D^*_1,\ldots,D^*_{J^*}|J^*,\chi,X,Z}(\cdot,\ldots,\cdot|\cdot,\ldots,\cdot)$ is rationalized given Assumption 1 as long as χ is a deterministic function of (θ,a,X,Z) as implied by the theoretical model.

To construct $F(\cdot,\cdot|\cdot,\cdot)$ we follow again the identification argument. Let $f(\theta|c,X,Z) = \kappa \tilde{f}_{\tilde{\theta}|\chi,X,Z}(\kappa\theta|c,X,Z)$ and $f(\theta|X,Z) = \kappa \tilde{f}_{\tilde{\theta}|X,Z}(\kappa\theta|X,Z)$, where these densities exist by condition (iii). In particular, $f(\theta|X,Z)$ is strictly positive on its support $[\underline{\tilde{\theta}}(1,x,z)/\kappa, \overline{\tilde{\theta}}(2,x,z)/\kappa] \subset I\!R_{++}$. Turning to $F_{a|\theta,X,Z}(\cdot|\cdot,\cdot,\cdot) = F_{a|\theta,X}(\cdot|\cdot,\cdot)$ by Assumption 2-(i), we follow (21). For every $(\theta,x) \in \mathcal{S}_{\theta X}$, let $F_{a|\theta,X}(\cdot|x,z)$ have a strictly positive density on its support $\mathcal{S}_{a|\tilde{\theta}x} \equiv \{a : \exists z \in \mathcal{S}_{Z|\tilde{\theta}x}, a = \tilde{a}(\tilde{\theta},x,z)\} = \mathcal{S}_{a|\theta x} \equiv \{a : \exists z \in \mathcal{S}_{Z|\theta x}, a = a(\theta,x,z)\}$ satisfying

$$F_{a|\theta,X}[a(\theta,x,z)|\theta,x] = \frac{\tilde{f}_{\tilde{\theta}|\chi,X,Z}(\tilde{\theta}|1,x,z)\psi(1|x,z)}{\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x,z)} \tag{A.1}$$

for every $(\theta,x,z) \in \mathcal{S}_{\theta XZ}$, where $\tilde{\theta} = \kappa\theta$ and $a(\theta,x,z) \equiv \tilde{a}(\kappa\theta,x,z)$. By (iv) the right-hand side has range $[0,1]$ as z varies in $\mathcal{S}_{Z|\tilde{\theta}x}$ for every given $(\tilde{\theta},x) \in \mathcal{S}_{\tilde{\theta}X}$, i.e. for every given $(\theta,x) \in \mathcal{S}_{\theta X}$. Thus, for every $(\theta,x) \in \mathcal{S}_{\theta X}$ and every $a \in \mathcal{S}_{a|\theta x}$, there exists a $z \in \mathcal{S}_Z$ such that $a = a(\theta,x,z)$, i.e. Assumption 2-(ii) is satisfied. We can now extend $F_{a|\theta,X}(\cdot|\theta,x)$ over $\mathcal{S}_{a|\theta x}$ by $F_{a|\theta,X}(a|\theta,x) = F_{a|\theta,X}[a(\theta,x,z)|\theta,x]$ using the above equation. Thus, $F(\cdot,\cdot|\cdot,\cdot) \in \mathcal{F}_{XZ}$ as desired.

The structure $[F(\cdot,\cdot|\cdot,\cdot), H(\cdot|\cdot,\cdot)]$ constructed as above rationalizes $\Psi_{J^*|\chi,X,Z}(\cdot|\cdot,\cdot,\cdot)$ because of (18) and the uniqueness of the corresponding density. This structure also rationalizes $\Psi_{\chi|X,Z}(\cdot|\cdot,\cdot)$. Specifically, by definition we have

$$F_{a|\theta,X}(a(\theta,x,z)|\theta,x) = \frac{f_{\theta|\chi,X,Z}(\theta|1,x,z)\nu_1(x,z)}{f_{\theta|X,Z}(\theta|x,z)} = \frac{\tilde{f}_{\tilde{\theta}|\chi,X,Z}(\tilde{\theta}|1,x,z)\nu_1(x,z)}{\tilde{f}_{\tilde{\theta}|X,Z}(\tilde{\theta}|x,z)}.$$

Using (A.1) shows that $\nu_1(x,z) = \psi_{\chi|X,Z}(1|x,z)$ as desired. The fact that the structure rationalizes $(t_1(\cdot,\cdot), dd_1(\cdot,\cdot), t_2(\cdot,\cdot), dd_2(\cdot,\cdot))$ follows the argument of the last paragraph of the proof of Lemma 3. □

References

Aryal, G. and **I. Perrigne** (2009): "An Insurance Model with Bidimensional Screening," Working Paper, Pennsylvania State University.

Aryal, G. and **I. Perrigne** (in progress): "Car Choice and Insurance Coverage," Working Paper, Pennsylvania State University.

Aryal, G., **I. Perrigne**, and **Q. Vuong** (2009): "Nonidentification of Insurance Models with Probability of Accidents," Working Paper, Pennsylvania State University.

Aryal, G., **I. Perrigne** and **Q. Vuong** (in progress): "An Empirical Analysis of Multidimensional Screening in Automobile Insurance," Working Paper, Pennsylvania State University.

Athey, S. and **P. Haile** (2007): "Nonparametric Approaches to Auctions," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume VI*, Amsterdam: North Holland.

Berry, S. and **P. Haile** (2009): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Working Paper, Yale University.

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

Brown, D. and **M. Wegkamp** (2002): "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica*, 70, 2035-2051.

Campo, S., **E. Guerre**, **I. Perrigne** and **Q. Vuong** (2009): "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders," *Review of Economic Studies*, forthcoming.

Carneiro, P., **K. Hansen** and **J. Heckman** (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361-422.

Chernozhukov, V., **H. Hong** and **E. Tamer** (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

Chiappori, P.A. and **B. Salanie** (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56-78.

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

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

D'Haultfoeuille, X. and **P. Fevrier** (2007): "Identification and Estimation of Incentive Problems: Adverse Selection," Working Paper, CREST Paris.

Einav, L., **A. Finkelstein** and **P. Schrimpf** (2009): "The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," *Econometrica*, forthcoming.

Gayle, G.L. and **R. Miller** (2008): "Identifying and Testing Generalized Moral Hazard Models of Managerial Compensation," Working Paper, Carnegie Mellon University.

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

Guerre, E., **I. Perrigne** and **Q. Vuong** (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193-1227.

Haile, P. and **E. Tamer** (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1-51.

Heckman, J.J. (2001): "Essays: Econometrics and Empirical Economics," *Journal of Econometrics*, 100, 3-5.

Hurwicz, L. (1950): "Generalization of the Concept of Identification," in T.C. Koopmans, ed., Statistical Inference in Dynamic Economic Models, Volume 10, Cowles Commission Monograph, New York: Wiley.

Kovchegov, Y. and **N. Yildiz** (2009): "Inference in Partially Identified Nonparametric Instrumental Variables Models," Working Paper, University of Rochester.

Koopmans, T.C. (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125-144.

Imbens, G. and **W. Newey** (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Wihtout Additivity," *Econometrica*, forthcoming.

Leslie, P. (2004): "Price Discrimination in Broadway Theatre," *Rand Journal of Economics*, 35, 520-541.

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145-177.

Lu, J. and **I. Perrigne** (2008): "Estimating Risk Aversion from Ascending and Sealed-Bid Auctions: The Case of Timber Auction Data," *Journal of Applied Econometrics*, 23, 871-896.

Lukacs, E. (1960): *Characteristic Functions*, London: Charles Griffin and Company Limited.

Manski, C. and **E. Tamer** (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

Matzkin, R. L. (1992): "Nonparametric and Distribution-Free Estimation of The Binary Threshold Crossing and The Binary Choice Models ," *Econometrica*, 60, 239-270.

Matzkin, R. L. (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58, 137-168.

Matzkin, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Volume IV*, North Holland.

Matzkin, R. L. (2007): "Nonparametric Identification," in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume VI*, North Holland.

Perrigne, I. and **Q. Vuong** (2008): "Nonparametric Identification of Contract Models With Adverse Selection and Moral Hazard," Working Paper, Pennsylvania State University.

Perrigne, I. and **Q. Vuong** (2009): "Nonlinear Pricing in Yellow Pages," Working Paper, Pennsylvania State University.

Pioner, H. (2007): "Semiparametric Identification of Multidimensional Screening Models," Working Paper, University of Chicago.

Rochet, J.C. and **P. Chone** (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 66, 783-826.

Rochet, J.C. and **L. Stole** (2003): "The Economics of Multidimensional Screening," in *Advances in Economic Theory: Eighth World Congress*, by Dewatripont, M., Hansen, L. P., and Turnovsky, S.J., eds, Cambridge University Press.

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

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

Su, L. and **H. White** (2008): "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory*, 24, 829-864.