Identification of Insurance Models with Multidimensional Screening

Gaurab Aryal
Pennsylvania State University

Isabelle Perrigne
Pennsylvania State University

Quang Vuong
Pennsylvania State University

July 2009

Incomplete Draft

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. Assuming constant absolute risk aversion, the model relies on the concept of certainty equivalence, while the joint distribution of risk and risk aversion is left unspecified. The object of our paper is to analyze how data availability constraints identification under four data scenarios from the ideal situation to a more realistic one. We consider insurance contracts defined by a premium and deductible. 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. The truncation arises from that the 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 more realistic 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 is not new and dates back from the twenties. See also Koopmans (1949) and Hurwicz (1950). The study of identification is a key element 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 what restrictions the model imposes on the observables as such restrictions 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 from this literature. 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 their willingness-to-pay distribution and their marginal utility. See Perrigne and Vuong (2009). In this respect, 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. 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. 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.[1]

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)

[1]Crawford and Shum (2007) consider two contracts but 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.

and Einav, Finkelstein and Schrimpf (2007) have shown an important heterogeneity in risk preferences, which may counterbalance the traditional intuition behind the Rothschild 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. Moreover, insurance contracts are of the form premium and deductible. 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 because it 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 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 data on 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 with a fully observed damage distribution. 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. 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), which solves 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 (i) using certainty equivalence is not suboptimal for the insurer to screen insurees and (ii) it is then optimal for the insurer to propose contracts of the form premium and deductible. 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, we show in a separate paper that the model is not identified even in the best data scenario of a continuum of contracts and a fully observed damage distribution. Intuitively, we 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. See Aryal, Perrigne and Vuong (2009). Because we exploit the observed number of accidents for each insuree, for conveneience 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 to the insuree only leading to a problem of imperfect information for the insurer. The latter is assumed to be a monopolist. 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 $F(\cdot,\cdot)$ on $\Theta \times \mathcal{A} = [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}]$, which is twice continuously differentiable on its support. 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,

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

We can verify that $\partial CE(0,0;\theta,a)/\partial\theta < 0$ and $\partial CE(0,0;\theta,a)/\partial a < 0$. The certainty equivalence of no insurance coverage decreases in both risk and risk aversion. We denote it by s. 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

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

The Insurer's Profit

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

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

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 not buying insurance $CE(0, 0; \theta, a) = s$. 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 certainty equivalence. The contracts need to guarantee the insuree's participation and his true type revelation. For the latter, we have

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

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

$$dd'(s) = -\eta(s, a, dd)t'(s), \quad \forall s \in [\underline{s}, \overline{s}], \forall a \in \mathcal{A}_s,$$

where $\mathcal{A}_s$ is the set of a values such that $CE(0, 0; \theta, a) = s$ for some $\theta \in \Theta$ and

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

with θ solution of $CE(0, 0; \theta, a) = s$, i.e. $\theta = a(w - s)/(\phi_a - 1)$. 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 set 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 $CE(0, 0; \underline{\theta}, \underline{a}) \equiv C_0(\overline{s})$.

The insurer maximizes his expected profit subject to the insuree's incentive compatibility and individual rationality constraints. The Hamiltonian is

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

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, $a^+(s)$ denotes the optimal path along which we solve the problem. This step can be viewed as a local optimization routine that allows us to solve the problem globally. Specifically, all the incentive compatibility constraints are satisfied as long as the incentive compatibility constraints for the $(s, a^+(s))$ insuree are satisfied. Moreover, by considering only those individuals, the insurer's expected profit is maximized. Formally, $a^+(s)$ maximizes $\mathrm{E}[\pi|s]$ with respect to a, i.e. $a^+(s)$ solves

$$\eta(s, a^+(s), dd(s))\mathrm{E}[\theta|s][1 - H(dd(s))] = 1.$$

Solving for the first-order conditions, $(t(s), dd(s))$ is solution of

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

where $\eta'(s, a^+(s), dd)$ denotes the total derivative of $\eta(s, a^+(s), dd)$ with respect to s, with the initial condition $CE(t(dd(\overline{s})), dd(\overline{s}); \overline{s}) = C_0(\overline{s})$. 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 needs to be viewed with C 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

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

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) optimally 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:

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

As shown by Aryal and Perrigne (2009), the only constraint that binds is the individual rationality constraint for the $(\underline{\theta}, \underline{a})$ insuree, i.e. $CE(t_1, dd_1; \underline{\theta}, \underline{a}) = C_0(\underline{\theta}, \underline{a}) \equiv CE(0, 0; \underline{\theta}, \underline{a})$. Maximizing $\mathrm{E}[\pi]$ with respect to (t_1, dd_1, t_2, dd_2) subject to this binding 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 individual rationality 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.[2] The problem of identification is to recover uniquely the structure $[F(\cdot,\cdot), H(\cdot)]$ 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 $(D_1, \ldots, D_J)$ possibly truncated at dd.

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.[3] 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. We will discuss this issue further in the conclusion within the framework of a semistructural model.

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

[2]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 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.

[3]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 in the model, we can consider that the relevant wealth is the value of the car. Einav and Cohen (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.

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 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 $\mathcal{F} \times \mathcal{H}$ as defined below.

Definition 1: *Let $\mathcal{F}_{XZ}$ be the set of conditional joint 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) (θ, a, J, X, Z) is i.i.d. across insurees,
(ii) $(D_1, \ldots, D_J) \perp (\theta, a)|(J, X, Z)$.
(iii) $(D_1, \ldots, D_J)|(J, X, Z)$ are i.i.d. as $H(\cdot|X, Z)$,
(iv) $J \perp (X, Z, a)|\theta$ with $J|\theta \sim \mathcal{P}(\theta)$, i.e. $\Pr[J = j] = e^{-\theta}\frac{\theta^j}{j!}$.

Assumption 1-(ii) says that conditional on the insuree's characteristics (X, Z), the amount of damage does not provide any information on his risk and risk type. 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-(iii) says that damages are independent conditional on (X, Z). Regarding Assumption 1-(iv), the number of accidents J depends on the insuree's risk θ, 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.

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.[4] We have $G(dd) = \Pr(\tilde{dd} \leq dd) = \Pr(\tilde{s} \leq s(dd)) = K(s)$ implying $g(dd) = k(s)s'(dd)$ with $s(\cdot)$ the inverse of $dd(\cdot)$ by monotonicity of the latter. Taking the ratio gives

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

[4]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 the 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).

Substituting the above expression in (5), we obtain

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

From (6), we have $t'_+(dd)=-1/\eta(s,a^+(s),dd)$, where $t_+(dd)=t[s(dd)]$. We also have $dt'_+(dd(s))/ds=-d[\eta(s,a^+(s),dd)]^{-1}/ds$, i.e. $t''_+(dd)\times dd'(s)=\eta'(s,a^+(s),dd)/[\eta(s,a^*(s),dd)]^2$ or equivalently $t''_+(dd)=[\eta'(s,a^+(s),dd)/[\eta(s,a^+(s),dd)]^2]\times s'(dd)$. Using this result, we can rewrite the previous equation as

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

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

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

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

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

showing that $a^+(s)$ is identified because all 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. $\mathrm{E}(\theta|dd)=\mathrm{E}(J|dd)$.[5] On the other hand, from the theoretical model, we know that $a^+(s)$ satisfies $\eta(s,a^+(s),dd(s))\mathrm{E}[\theta|s][1-H(dd(s))]=1$. Using (1) and (4), we obtain

$$s=w-\frac{\mathrm{E}[\theta|dd]}{a^+(s)e^{a^+(s)dd}}\left(\mathrm{E}[e^{a^+(s)D}]-1\right),$$

[5]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-(iv) and the one-to-one mapping between (θ,a) and (θ,s).

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 claims are observed. Under Assumption 1, the structure* $(K(\cdot), H(\cdot))$ *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 crucial role in identifying $F(\cdot,\cdot)$.[6] Specifically, with claim data we construct the moment generating function of the number of accidents J conditional on s, which identifies the moment generating function of θ given s in a neighborhood of zero. As is well known, this 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 $(t(s), dd(s))$ with 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-(iv) 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

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

[6] 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).

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).[7]

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 $T(\theta, a) = [\theta, w - [\theta(\phi_a - 1)]/a]'$ 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 claims are observed. 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 abstration of dynamic considerations, an accident leads to a claim if and only if the damage is above the deductible. Using the claim data, 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 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-(ii). Thus, we have the following lemma.

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

Once $H(\cdot)$ is identified, it remains to study the identification of $F(\cdot, \cdot)$. Intuitively the argument is the same as in Case 1, though the reported number of accidents J^* is

[7] 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.

observed instead of J. Reviewing the argument leading to Lemma 1, it is straightforward to see that $K(\cdot)$ is identified if $\mathrm{E}(\theta|dd)$ is identified. 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-(ii,iii). Thus, $\mathrm{E}[J^*|dd] = \mathrm{E}\{\mathrm{E}[J^*|J, dd]|dd\} = \mathrm{E}[J(1-H(dd))|dd] = (1-H(dd))\mathrm{E}[J|dd] = (1 - H(dd))\mathrm{E}(\theta|dd)$. Hence, $\mathrm{E}[\theta|dd]$ is identified despite the truncation of the damage distribution at dd leading to the identification of $K(\cdot)$.

Regarding 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

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

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 claims 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 need to exploit sufficient variations in exogenous variables to achieve identification. Moreover, 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. 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. Their robustness analysis suggests that when considering optimality of the offered contracts this correlation turns out to be negative as expected. Second, by assuming that the number of claims follows a Poisson distribution to recover the marginal distribution of risk, they face the classical problem of overdispersion, i.e. the empirical variance tends to be too large relative to the sanple mean thereby contradicting the Poisson distribution. Our identification results rely instead on a nonparametric mixture of Poisson distribution for the number of claims thereby avoiding the overdispersion problem. 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, to achieve identification of the distribution of a given (θ, X, Z) on its support, we make an exclusion restriction and a full support assumption involving the car characteristics Z.

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 provide 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, 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 $u \in (-1, +\infty)$.[8] 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 all 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 $\mathcal{S}_X$, $\mathcal{S}_Z(x)$, $\Theta(x)$ and $\mathcal{A}(x)$ be the supports of X, $Z|X = x$, $\theta|X = x$ and $a|X = x$, respectively.

Assumption 2: *We have*

(i) $a \perp Z|(\theta, X)$

(ii) $\forall x \in \mathcal{S}_X$ *and* $\forall(\theta, a) \in \Theta(x) \times \mathcal{A}(x)$, *there exists* $z \in \mathcal{S}_Z(x)$ *such that* $a(\theta, x, z) = a$.

[8] 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 5.1.

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 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 value 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.[9] 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 claims are observed. 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 to each insuree and claims observed

[9] 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.

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 claim data are available and there is enough variation in some excluded exogenous variables. The truncation on the damage distribution, however, will limit the extent of our identification result. Nevertheless, we show that $F(\cdot,\cdot)$ is identified up to the knowledge of $H(dd_2(X,Z)|X,Z)$ or equivalently $H(dd_1(X,Z)|X,Z)$, where $dd_1(X,Z) > dd_2(X,Z)$.[10]

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 equation shows

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

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

$$M_{J^*|\chi,X,Z}(t|c,x,z) \;=\; \mathrm{E}[e^{J^* t}|\chi = c, X = x, Z = z]$$

[10] 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 $H_2(\cdot|X,Z)$.

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

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

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

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

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

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

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

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

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

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

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

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 claims are observed only when the damage is 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

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

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

In view of of Proposition 4, the identification of $H_2(X,Z)$ would lead to identify the structure.[11] 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).

Lemma 3: *Suppose that two insurance coverages are offered to each insuree and claims are observed only when the damage is 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

to be completed

[11] 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.

6 Identification Strategies for Case 4

From Section 4.2, it is clear that any identifying assumption that pins down $H_2(X, Z)$ is necessary and sufficient to 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 $H_2(X, Z)$ would be 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 $H_2(x, z)$.

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

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

where $\mathrm{E}[D|D \geq dd_2(x,z)|X = x, z = z] = \int_{dd_2(x,z)}^{\overline{d}(x,z)} Dh_2^*(D|X = x, Z = z)dD$ 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 $\mathrm{E}[D|D \leq dd_2(x,z)|X = x, Z = z]$ 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 $\mathrm{E}[D|D \leq dd_2(x_0, z_0)|X = x_0, Z = z_0]$ 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 $H_2^0(x,z)$.[12] 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

$$\left(1 - \frac{1 - H_2^0(x,z)}{\sup_{(\tilde{x},\tilde{z})}[1 - H_2^0(\tilde{x},\tilde{z})]}, 1\right). \tag{18}$$

For the values (x,z) for which $1 - H_2(x,z)$ 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*

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

for every $D \leq dd_2(x,z)$ 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 $0 \leq H_2(x,z) \leq dd_2(x,z)h(dd_2(x,z)|x,z)$. Dividing both sides by $1 - H_2(x,z)$ and using the definition of the conditional density $h_2^*(\cdot|x,z)$, we obtain

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

Solving for $H_2(x,z)$ gives

$$0 \leq H_2(x,z) \leq \frac{dd_2(x,z)h_2^*(dd_2(x,z)|x,z)}{1 + dd_2(x,z)h_2^*(dd_2(x,z)|x,z)} \equiv \overline{B}(x,z). \tag{19}$$

[12] 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 $[F^0(\cdot,\cdot|X,Z), H^0(\cdot|X,Z)]$.

Similarly, exploiting the relationship $1 - H_2(x,z) = [1 - H_1(x,z)]/\pi(x,z)$ we obtain

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

We note that the upper bound for both $H_1(x,z)$ and $H_2(x,z)$ is strictly less than 1 and the lower bound for $H_1(x,z)$ is strictly larger than zero, thereby thightening the bounds. A main advantage of these bounds is that they are expressed as functions of observables.[13]

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 $(x,z) \in \mathcal{S}_{XZ}$.

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 identified c.d.f. $\tilde{F}_{\tilde{\theta}|X,Z}(\cdot|x,z)$.

To obtain such an explicit form, we consider its density $\tilde{f}_{\tilde{\theta}|X,Z}(\cdot|x,z)$. 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 $\phi_{\tilde{\theta}|\chi,X,Z}(\cdot|c,x,z) = \mathrm{E}[e^{i\zeta\tilde{\theta}}|\chi = c, X = x, Z = z]$ 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 $\mathcal{B}(n,p)$

[13] To show that these bounds are sharp, it would require 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 this bound is between 0 and the lower bound of the interval (18). Moreover, the lower bound of (18) is expressed in terms of the true value $H_2^0(x,z)$, which is not identified.

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

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

where the second equality follows from $\tilde{\theta}=(1-H_2(X,Z))\theta$. Hence, we have

$$\phi_{\tilde{\theta}|\chi,X,Z}(\tilde{\zeta}|c,x,z)=\phi_{J^*|\chi,X,Z}(\zeta|c,x,z), \quad \text{where } \tilde{\zeta}=\frac{1-H_c(x,z)}{1-H_2(x,z)}\frac{e^{i\zeta}-1}{i}, \tag{20}$$

for all $\zeta \in \mathbb{C}$. Let $\zeta = u + i\log(\cos u)$ 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 $c=1$, 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

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

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

to be completed

Appendix

Proof of Lemma 3: In view of Proposition 4, $H_2(X,Z)$ is identified if and only if the structure $[F(\cdot,\cdot|X,Z), H(\cdot|X,Z)$. Thus, it usffices 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 $(\chi, J^*, D_1^*, \ldots, D_{J^*}^*, t_1, dd_1, t_2, dd_2)$ 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 $\tilde{\theta}(a,X,Z)=\kappa\theta(a,X,Z)$, $\tilde{f}(\tilde{\theta}(a),a|X,Z)=f(\tilde{\theta}(a)/\kappa,a|X,Z)/\kappa=f(\theta(a),a|X,Z)/\kappa$, $1-\tilde{H}(D|X,Z)=(1-H(D|X,Z))/\kappa$, $\tilde{\nu}_c=\nu_c$ 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.$\square$

References

Aryal, G. and **I. Perrigne** (2009): "An Insurance Model with Bidimensional Screening," 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.

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.

Campo, S., **E. Guerre**, **I. Perrigne** and **Q. Vuong** (2009): "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders," Working Paper, Pennsylvania State University.

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** (2007): "The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market," Working Paper, Stanford University.

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*, forthcoming.

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.

Rotschild, 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.