WORKING PAPERS IN ECONOMICS & ECONOMETRICS

Nonidentification of Insurance Models with Probability of Accidents

Gaurab Aryal
Research School of Economics
College of Business and Economics
Australian National University
E-mail: gaurab.aryal@anu.edu.au

Isabelle Perrigne
Pennsylvania State University

Quang Vuong
Pennsylvania State University

JEL codes: C14, L62, D82, D86

Working Paper No: 552
ISBN: 0 86831 552 4

August, 2011

Abstract

In contrast to Aryal, Perrigne and Vuong (2009), this note shows that in an insurance model with multidimensional screening when only information on whether the insuree has been involved in some accident is available, the joint distribution of risk and risk aversion is not identified.

Nonidentification of Insurance Models with Probability of Accidents

G. Aryal, I. Perrigne & Q. Vuong

1 Introduction

This note studies the nonparametric identification of the joint distribution of risk and risk aversion where data contain information on whether an insuree have had involved in an accident.[1] Aryal and Perrigne (2010) characterizes the optimal insurance contracts sold by an insurer when insurees have private information about their risk and risk aversion. Under the constant absolute risk aversion assumption, the paper shows that the certainty equivalence without insurance coverage is a one dimensional sufficient statistics that effectively reduces the two dimensional private information into one. Identification of the distribution of certainty equivalence follows the same logic as in identification of distribution of private value in first price auction, see Guerre, Perrigne and Vuong (2000). The analogous of bids here is the (observed) choice of deductible, the unobserved private valuation is the certainty equivalence and the one to one mapping between the two is provided by the first order conditions that characterize optimal coverage. Although the distribution of certainty equivalence is identified the joint distribution of risk and risk aversion cannot be nonparametrically identified. Thus, this note complements Aryal, Perrigne and Vuong (2009), where the risk is defined as the expected number of accidents and the model is nonparametrically identified.

[1] In both theoretical research starting with Rothschild and Stiglitz (1976); as well as empirical research starting with Chiappori and Salanie (2001) on insurance risk has always been interpreted as the probability of accident.

The identification is exactly the same as in Aryal, Perrigne and Vuong (2009) in Case 1 with the risk being interpreted as probability of having an accident. This identification result allows us to recover (pseudo) certainty equivalence for every deductible we observe in the data. The second step is then to use this information on the number of accidents to recover the conditional distribution of risk for given certainty equivalence. Conditional on a particular coverage (and hence certainty equivalence) the number of accidents is only a function of risk and not risk aversion, and provides information on conditional all the moments of risk and therefore the distribution. However, when we interpret risk as the probability accident, the only information we can use is whether or not an insuree have been in an accident and not the number of accidents: insurees with one or more than one accident claims are treated as the same, which eliminates the variation in observed claims to differentiate the riskiness of insurees. This variation in claims data is very important for identification. We also introduce some exogenous variation in insuree and car characteristics to explore the possibility of identification. We show that even under some strong exclusion restriction assumption, in particular independence between the exogenous characteristics and risk and risk aversion, the joint density function is not identified. The non identification result relies on the characterization of identification of arbitrary mixtures.

The rest of the paper is organized as follows. Section 2 briefly outlines the model and introduces. Section 3 presents the main result of the paper: namely, First the distribution of the certainty equivalence is identified. Second, the joint distribution of risk and risk aversion is not identified from knowledge of distribution of certainty equivalence. Third, it is shown that even with arbitrary variation in exogenously observed variable and under any relevant exclusion restriction the model is not identified.

2 The Model

The aim of this section is to introduce the notations and the model. For a more detailed analysis see Aryal and Perrigne (2010). An insuree is characterized by a risk θ, which is the probability of accident and a CARA coefficient a. Thus the utility function is $U_a(x) = -e^{-ax}, a > 0$. The pair (θ, a) is distributed as $F(\cdot,\cdot)$ on $\Theta \times \mathcal{A} = [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}]$. When there is an accident an it incurs a damage D. The damage is modeled as a random variable distributed as $H(\cdot)$ on $[0, \overline{d}]$. It is assumed that the damage D is independent of (θ, a).[2]

Certainty Equivalence

Let, $w > 0$ be the weath of an insuree which is observed. When an insuree of type (θ, a) buys no coverage his expected utility is

$$V(0,0;\theta,a) = -e^{aw}\left[(1-\theta) + \theta \int_0^{\overline{d}} e^{aD} \mathrm{d}H(D)\right].$$

Certainty equivalence ($CE(0,0;\theta,a)$) is defined as the certain amount which makes the insuree as well off as without any coverage. More specifically, let $s = CE(0,0;\theta,a)$ then $-e^{aCE(0,0;\theta,a)} = -e^{aw}\left[(1-\theta) + \theta \int_0^{\overline{d}} e^{aD} \mathrm{d}H(D)\right]$,

$$s = w - \frac{\log\left[1 + \theta\left\{\int_0^{\overline{d}} \exp(aD)\mathrm{d}H(D) - 1\right\}\right]}{a}.$$

Since s is a function of (θ, a), it is also a random variable distributed as $K(\cdot)$[3]. More specifically,

$$K(\tilde{s}) = \Pr(s \leq \tilde{s}) = 1 - \int_{\underline{a}}^{\overline{a}} \int_{\underline{\theta}}^{\tau(a,\tilde{s})} \mathrm{d}F(\theta, a),$$

where $\tau(a,\tilde{s}) = \frac{\exp(a(w-\tilde{s}))-1}{\int_0^{\overline{d}} \exp(aD)\mathrm{d}H(D)-1}$. Let, (t, dd) pair denote a coverage where t is the premium a dd the deductible. An insuree of type (θ, a) then chooses the coverage (t, dd) that maximizes his/her expected utility

$$V(t,dd;\theta,a) = (1-\theta)u_a(w-t) + \theta\left[\int_0^{dd} u_a(w-t-y)\mathrm{d}H(y) + u_a(w-t-dd)\left(1-H(dd)\right)\right],$$

[2]In other words, knowing D does not carry any information on the risk and/or the risk aversion.

[3]Since a larger risk and/or risk aversion results in lower certainty equivalence without a coverage, $\underline{s}$ ($\overline{s}$) corresponds to the certainty equivalence of $(\overline{\theta}, \overline{a})$ $((\underline{\theta}, \underline{a}))$, respectively.

which is equivalent to choosing (t, dd) to maximize the corresponding certainty equivalence

$$CE(t, dd; \theta, a) = w - t - \frac{\log[\int_0^{dd} e^{aD} \mathrm{d}H(D) + e^{add}(1 - H(dd)) - 1]}{a}.$$

Insurer's Profit

A risk neutral insurer offers a contract/coverage (t, dd) that maximize its expected profit:

$$\mathrm{E}(\pi) = \int_{\underline{a}}^{\overline{a}} \int_{\underline{\theta}}^{\overline{\theta}} \left\{ t(\theta, a) - \theta \left[\int_0^{\overline{d}} \max\{0, D - dd(\theta, a)\} \mathrm{d}H(D) \right] \right\} \mathrm{d}F(\theta, a)$$

subject to incentive compatibility and individual rationality constraints:

(IC): $\quad V(t(\theta, a), dd(\theta, a), \theta, a) \geq V(t(\widetilde{\theta}, \widetilde{a}), dd(\widetilde{\theta}, \widetilde{a}), \theta, a), \forall (\theta, a), (\widetilde{\theta}, \widetilde{a}) \in \Theta^2 \times \mathcal{A}^2.$

(IR): $\quad V(t(\theta, a), dd(\theta, a), \theta, a) \geq V(0, 0, \theta, a), \forall (\theta, a) \in \Theta \times \mathcal{A}.$

Following Aryal and Perrigne (2010), make the change of variable from (θ, a) to (θ, s) and note that $t(\theta, s) \equiv t(s)$, and after some simplification the expected profit is :

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

Optimization Problem

The objective of the insurer is to design contract $(t(s), dd(s))$ such that it maximizes the expected profit subject to appropriate (IC) and (IR), which can be written in terms of certainty equivalence. Aryal and Perrigne (2010) show that it is enough to ensure that the (IR) binds for insurees of type $\underline{s}$. The (IC) constraints imply that the insuree will report their certainty equivalence to be $\tilde{s}$ that maximizes his/her certainty equivalence from the coverage corresponding to the reported $\tilde{s}$. The local (IC) is then given by $\max_{\tilde{s} \in [\underline{s}, \overline{s}]} CE(t(\tilde{s}), dd(\overline{s}); \theta, a)$ where at $s = \tilde{s}$

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

and $\eta(s,a,dd) = \frac{1}{\theta e^{add}(1-H(dd))}$. Formulating the appropriate Hamiltonian, the optimal contract is characterized by the solutions of the following two FOCs

$$\eta(s,a(s))\mathrm{E}(\theta|s)(1-H(dd)) \\ +\frac{K(s)}{k(s)}\frac{1}{\eta(s)}\left[-\frac{\partial\eta(s,a(s),dd(s))}{\partial dd}\frac{\mathrm{d}(dd)}{\mathrm{ds}}+\eta'(s,a(s))\right]=1 \tag{1}$$

$$\frac{\mathrm{d}dd(s)}{\mathrm{ds}} = -\eta(s,a(s))\frac{\mathrm{d}t(s)}{\mathrm{ds}} \tag{2}$$

with the following boundary conditions:

$$dd(\underline{s})=0; \quad t(\overline{s}) = \frac{1}{\underline{a}}\log\left[\frac{1+\underline{\theta}\{\int_0^{\overline{d}} e^{ay}\mathrm{d}H(y)-1\}}{1-\underline{\theta}}\right].$$

3 Identification

The model structure is defined as $F(\cdot,\cdot)$ and $H(\cdot)$. For every insuree i, we observe the coverage choice (t_i, dd_i); the variable $\chi_i \in \{0,1\}$ where $\chi_i = 1$ if there is an accident and 0 otherwise; the total amount of damage filed D. We also observe individual and car characteristics X and Z, respectively, where $(X,Z) \subset \mathbb{R}^{\mathrm{dimX}} \times \mathbb{R}^{\mathrm{dimZ}}$. Conditional on observed characteristics, $Z=z$ and $X=x$, the risk and risk aversion is distributed as $F(\cdot,\cdot|Z=z,X=x)$, on the set $[\underline{\theta}(x,z),\overline{\theta}(x,z)]\times[\underline{a}(x,z),\overline{a}(x,z)] \equiv \mathcal{S}_{\theta a|X,Z}$ and the damage $D \sim H(\cdot|x,z)$ on $[0,\overline{d}(x,z)]$. This model is then said to be identified if we can recover uniquely the structure $[F(\cdot,\cdot|X,Z),H(\cdot|X,Z)] \in \mathcal{F}_{XZ}\times\mathcal{H}_{XZ}$ from the observables. The definition of the admissible set of structures, $\mathcal{F}_{XZ}\times\mathcal{H}_{XZ}$ corresponds to Definition 1 and 2, respectively in Aryal, Perrigne and Vuong (2009).

Following Aryal, Perrigne and Vuong (2009) we have the following assumption:
Assumption 1

(i) (θ, a, χ, X, Z) is i.i.d across insurees.

(ii) $D \perp (\theta, a)\big|(\chi, X, Z)$.

(iii) D i.i.d as $H(\cdot|X, Z)$.

(iv) $\chi \perp (X, Z, a)\big|\theta$ with $\chi|\theta \sim B(\theta)$.

Assumption 1-(iv) tells us that for any insuree with risk θ, the event accident or no accident is distributed as a Bernoulli random variable with parameter θ as $\Pr[\chi = 1] = \theta$.

Identification of $K(\cdot)$ and $H(\cdot)$

After suppressing the dependence on (X, Z), we follow Case 1 in Aryal, Perrigne and Vuong (2009) to conclude that the structure $[K(\cdot), H(\cdot)]$ is identified. With complete information on the damages $H(\cdot)$ can be identified therefrom. Observe that the the FOCs characterizing optimal coverages as a function of s is the same as in Aryal, Perrigne and Vuong (2009) with only the interpretation of θ being probability of accident. Following the exact steps as in Aryal, Perrigne and Vuong (2009), we can also identify $a(s)$ using the FOCs (2), (3) and expressing them as a function of observables, using $\mathrm{E}(\theta|s) = \mathrm{E}(\theta|dd) = \mathrm{E}(\chi|dd)$, which is observed in the data. Then since $a(s)$ is identified, we can express s as a function of observables, using the definition of s and the (IC) to identify its distribution $K(\cdot)$. This result is then formalized as follows:

Proposition 1 (Aryal, Perrigne and Vuong (2009)): *Suppose 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.*

Nonidentification of $F(\cdot, \cdot)$

In this section, we show that $F(\cdot, \cdot)$ cannot be uniquely recovered from $[K(\cdot), H(\cdot)]$. Intuitively from the definition of s, the knowledge of $\Pr[s \leq \tilde{s}]$, determines $\Pr[R = (\theta, a) : CE(0, 0; \theta, a) \leq \tilde{s}]$. However this is not enough to determine the probability assigned to all the open sets in (rectangles) in $\mathbb{R}^2_{++}$. As an illustration, consider a case where (θ, a) can take finite values, with the same conditional mean $\mathrm{E}(\theta|s)$ but with different joint mass function, is provided below (see Fig.1). Case 1 and Case 2 corresponds to two joint prob-

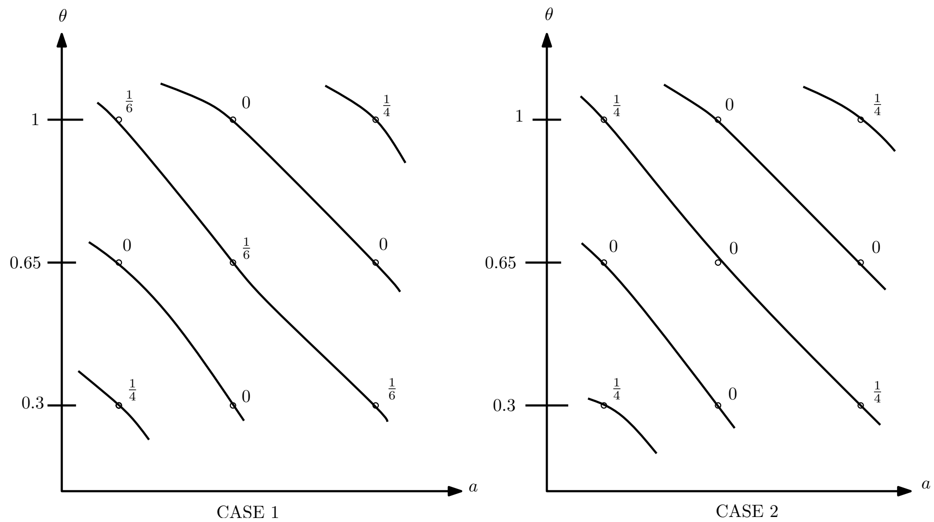

Figure 1: Finite Type Space

ability mass function which are observationally equivalent.[4] At the lowest risk $\underline{\theta} = 0.3$, there can be only one value of $\overline{s}$ for the lowest risk aversion $\underline{a}$, hence in Case 2, the value probability mass has to be the same, i.e. 1/4. Similarly for $\overline{\theta}$ and $\overline{a}$. The most interesting thing is the third iso-certainty equivalence curve from the origin. There are three risk and risk aversion pairs that correspond to the same certainty equivalent. Both cases 1 and 2 the conditional mean of θ is the same, hence are observationally equivalent although the conditional mass functions are different. This example shows that $\Pr[\tilde{\theta} = \theta|s]$ is a function of only first moment of $f_{\theta|S}(\theta|s)$ and can be extended to the case with continuous (θ, a) and is formalized below:

Proposition 3.1. *Suppose a continuum of insurance coverages is offered to each insuree and all claims are observed. Under Assumption 1', $F(\cdot,\cdot)$ is not identified.*

Proof. Let $f_{\theta,a}(\cdot,\cdot)$ and $\tilde{f}_{\theta,a}(\cdot,\cdot)$ be two joint density functions of (θ, a). Then, because $f_{\theta,a}(\theta, a) = f_{\theta,s}(\theta, \lambda^{-1}(s;\theta))J$ where $f_{(\theta,s)}(\cdot,\cdot)$ is the joint density of $(\theta, s), \lambda(\theta, a) = s$ with J being the appropriate Jacobian of the transformation and similarly $\tilde{f}_{\theta,a}(\theta, a) =$

[4] As it can be seen, the level a can take is not important because we can choose a such that for any θ, the pair (θ, a) corresponds to a known s, i.e. $\lambda(\theta, a) = s$.

$\tilde{f}_{\theta,s}(\theta,\lambda^{-1}(s;\theta))J$, we know that $f_{\theta,a}(\cdot,\cdot)=\tilde{f}_{\theta,a}(\cdot,\cdot)$ if and only if $f_{\theta|s}(\cdot|\cdot)=\tilde{f}_{\theta|s}(\cdot|\cdot)$ because $k(s)$ is identified. We know $\Pr(\chi=1|s)$ from the data and can be expressed as a mixture:

$$\Pr(\chi=1|s) \;=\; \int_{A_s(\theta)} \Pr(\chi=1|\theta)f(\theta|s)\mathrm{d}\theta = \int_{A_s(\theta)} \theta f(\theta|s)\mathrm{d}\theta,$$

where $A_s(\theta)=\{\theta:\exists a, s=\lambda(\theta,a)\}$. Similarly, we have $\Pr(\chi=1|s)=\int_{A_s(\theta)}\theta\tilde{f}(\theta|s)\mathrm{d}\theta$, thus as long as $k(\theta|s)$ and $\tilde{k}(\theta|S)$ have the same first moments, they cannot be distinguished by the model and are therefore observationally equivalent.[5] □

As a simple example of nonidentification consider two joint distribution of risk and certainty equivalence: $f_{\theta S}(\cdot,\cdot)$ and $\tilde{f}_{\theta S}(\cdot,\cdot)$ such that

$$f_{\theta|S}(\cdot|s)=N(\mu,\sigma(s)); \quad \tilde{f}_{\theta|S}(\cdot|s)=N(\mu,\tilde{\sigma}(s))$$

and $\sigma(s)\neq\tilde{\sigma}(s),\forall s$, where $N(\cdot,\cdot)$ is the Normal density. Then,

$$f_{\theta,S}(\theta,s)=N(\mu,\sigma(s))\times k(s) \quad \& \quad \tilde{f}_{\theta,S}(\theta,s)=N(\mu,\tilde{\sigma}(s))\times k(s)$$

are observationally equivalent.

Nonidentification under Exclusion Restriction

Thus far we have not used the fact that the offered coverage vary with observed characteristic of the insuree and his/her car. In view of the result above, answer to whether or not the model is identified by using the variation in the observed covariates $(X,Z)\in\mathcal{S}_{XZ}$ under appropriate exclusion restriction.[6] However, even with arbitrary variation in (X,Z) and under the strongest exclusion restriction assumption, the model is not identified. The model is first interpreted as an arbitrary mixture model, which simplifies the exposition by making the analysis tractable and simple. The identification of the model is then identification of an appropriate mixture model. The objective will be to study identification

[5] More generally, it can be shown that a mixture of Binomial random variable with fixed n but variable p is not identifiable and Bernoulli random variable is a Binomial random variable with $n=1$.

[6] For instance, Guerre, Perrigne and Vuong (2009) use exclusion restriction in the form of exogenous entry in auction to nonparametrically identify the utility function, which otherwise could not be nonparametrically identified.

under the most strongest exclusion restriction assumption, i.e. $(\theta, a) \perp (X, Z)$, which makes the identification most likely.

Assumption 2': We have

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

Assumption 2' then implies $\mathcal{S}_{\theta a|X,Z} = \mathcal{S}_{\theta a}$ and $\mathcal{F}_{XZ} = \mathcal{F}$. For notational convenience we treat Z to be the only observed co-varaites.[7] We begin with the functional form of certainty equivalence, which is also the structural equation of the model

$$s = w - \frac{\log\left[1 + \left(\int_0^{\bar{d}(z)} e^{aD} h(D|z)\mathrm{d}D - 1\right)\theta\right]}{a} = w - \mathbf{v}(\theta, a; z).$$

Since we know s and w we know the distribution of $v(\theta, a; z)$ on the support $\mathcal{S}_{V|Z} \equiv \{v : \exists z \in \mathcal{S}_Z \;\; v = \mathbf{v}(\theta, a; z)\}$ for some $(\theta, a) \in \mathcal{S}_{\theta a}$. Let $\mathcal{Q}_Z$ be the space of all conditional density functions $q_{\mathbf{v}|Z}(\cdot|\cdot)$ conditional on Z, defined over $V \times Z$ with $Q_{\mathbf{v}|Z}(\cdot|\cdot)$ the corresponding distribution. Then, we can identify $F(\cdot, \cdot)$ from $K(\cdot)$ if and only if we can identify it from $Q_{\mathbf{v}|Z}(\cdot|\cdot)$, and hence the latter can be treated as the (structural) equation of interest. Thus we are interested in a model for a continuous outcome v that is defined by the following restrictions:

(Ra): For every $z \in \mathcal{S}_Z, v = \mathbf{v}(\theta, a; z)$ is continuously differentiable and continuously distributed with $q_{\mathbf{v}|Z}(\cdot|\cdot) \in \mathcal{Q}_Z$, monotonic (increasing) in θ and a.

(Rb): For every $z \in \mathcal{S}_Z, F_{\theta a|Z}(\theta, a|z) = F(\theta, a)$ and $\mathrm{E}(\theta|z) = \theta_z$ and θ_z is known and $\mathcal{F}$ is the set of all feasible bivariate distribution function, absolutely continuous with respect to Lebesgue measure and full support.

We observe a $i.i.d$ samples of $\{v_i, z_i\}_{i=1}^N$ and with sufficient observations i.e.$(N \to \infty)$, one can identify $Q_{\mathbf{v}|Z}(v|z)$, for any $v \in \mathcal{S}_{V|Z}$ and $z \in \mathcal{S}_Z$. Thus we have

$$Q_{\mathbf{v}|Z}(v|z) = \Pr[(\theta, a) \in \mathcal{S}_{\theta a} : \mathbf{v}(\theta, a; z) \le v] = F(\{(\theta, a) \in \mathcal{S}_{\theta a} | \mathbf{v}(\theta, a; z) \le v\}). \quad (3)$$

Let $(\mathcal{S}_{\theta a}, \mathcal{B}_{\theta a}, F)$ be a measurable space where $\mathcal{B}_{\theta a}$ includes singletons and $\phi : \mathcal{F} \to \mathcal{Q}_Z$

[7]This is without loss of generality because the argument neither distinguishes between X and Z nor relies on their dimension. Moreover, Z is more likely to be continuous and hence is the best hope to aid in identification.

defined as $\phi(F) = q_{\mathbf{v}|Z}(\cdot|\cdot)$ Then, $\phi(\cdot)$ and (3) are related by the following mixture:

$$q_{\mathbf{v}|Z}(v|z) = \int_{\mathcal{S}_{\theta a}} \delta(v, z; \theta, a)\mathrm{d}F(\theta, a),$$

where $\delta(v, z; \theta, a) = 1$ if $\mathbf{v}(\theta, a, z) = v$ and 0 otherwise, also known as the kernel and $F(\cdot, \cdot)$ is the mixing distribution. Hence, $\mathcal{Q}_Z$ is said to be identifiable if $\phi(\cdot)$ is one-to-one i.e. if $F, F' \in \mathcal{F}, F \neq F'$, then there exists in the data (v, z) such that $q_{\mathbf{v}|Z}(v|z) \neq q'_{\mathbf{v}|Z}(v|z)$. In other words, $F(\cdot)$ is identifiable if and only if ϕ is invertible, whence $F(\cdot, \cdot) = \phi^{-1}(q_{\mathbf{v}|Z})(v|z)$.

Let $\mathcal{L} = \{\delta(v, z; \theta, a)|\theta, a \in \mathcal{S}_{\theta a}\}$ be a family of atomic distributions which is measurable in $Z \times \mathcal{S}_{\theta a}$ indexed by $\theta, a \in \mathcal{B}_{\theta a}$. Let $\mathcal{L}_+ = \{\delta(v, z; \cdot)|(v, z) \in \mathcal{S}_{V|Z} \times \mathcal{S}_Z\}$. Let $C_0(\mathcal{S}_{\theta a})$ be the Banach space of continuous functions on $\mathcal{S}_{\theta a}$ that vanishes as infinity and the norm is $||p|| = \sup_{(\theta,a)\in\mathcal{S}_{\theta a}} |p(\theta, a)|$ for $p \in C_0(\mathcal{S}_{\theta a})$. Then we have the following characterization of the identifiability of mixture by Blum and Susarla (1977):

Theorem 1 *Suppose $\mathcal{L}_+ \subset C_0(\mathcal{S}_{\theta a})$. Then for $(v, z) \in \mathcal{S}_{V|Z} \times \mathcal{S}_V, q^1_{v|Z}(v|z) = q^2_{v|Z}(v|z) \Rightarrow F_1(\theta, a) = F_2(\theta, a), (\theta, a) \in \mathcal{S}_{\theta a}$ iff $\mathcal{L}_+$ is dense in $C_0(\mathcal{S}_{\theta a})$.*

Thus the linear space spanned by $\mathcal{L}_+$ being dense in $C_0(\mathcal{S}_{\theta a})$ is necessary and sufficient for the mapping $\phi(\cdot)$ to be bijective and hence identifiable (invertible). We further know that an easier characterization of a dense subset is given by the following result, see Conway (1985) Corollary III.6.14.

Lemma 1: *$< \mathcal{L}_+ >$ is dense in $C_0(\mathcal{S}_{\theta a})$ for $F(\cdot, \cdot) \in \mathcal{F}$ if and only if for all $\delta \in \mathcal{L}_+ \int \tilde{p}(\theta, a)\delta(v, z; \theta, a)\mathrm{d}F(\theta, a) = 0 \Rightarrow \tilde{p} \equiv 0, F(\cdot)-a.e.$ where $\tilde{p} \in L^\infty(F)$.*

The implication of this on identification is then immediate. Suppose $F_1 \neq F_2$ but $\int_{\mathcal{S}_{\theta a}} \delta(v, z; \theta, a)\mathrm{d}F_1(\theta, a) - \int_{\mathcal{S}_{\theta a}} \delta(v, z; \theta, a)\mathrm{d}F_2(\theta, a)$=0, then we have

$$\int_{\mathcal{S}_{\theta a}} \delta(v, z; \theta, a)\frac{(f_1(\theta, a) - f_2(\theta, a))}{f_1(\theta, a)}\mathrm{d}F_1(\theta, a) = 0 \Rightarrow \int_{\mathcal{S}_{\theta a}} \delta(v, z; \theta, a) \underbrace{\tilde{p}(\theta, a)}_{\in L^\infty(\mathcal{S}_{\theta a})} \mathrm{d}F_1(\theta, a) = 0.$$

Then from the Lemma 1 we have $\tilde{p}(\theta, a) = 0 - a.e.F_1$. But since $f_1(\cdot, \cdot) > 0$, we get $F_1 = F_2$, a contradiction. Let $\tilde{p}(\theta, a) = 1$ for all $(\theta, a \in)\mathcal{S}_{\theta a}$. Then for a fixed (θ, a), $\delta(\cdot, \cdot; \theta, a)$ puts positive weight (=1) only at that point $(\theta, a) \in \mathcal{B}_{\theta a}$. But because $F(\cdot, \cdot)$ is absolutely continuous, we get $\int \delta(v,z;\theta,a)\tilde{p}(\theta,a)\mathrm{d}F(\theta,a) = 0$ for all $\delta(\cdot,\cdot;y) \in \mathcal{L}_+$, hence $\mathcal{L}_+$ is not dense in $C_0(\mathcal{S}_{\theta a})$. As an example, let's consider a simple $\mathcal{B}_{(\theta,a)}$- measurable function that equal to 1 on $(\theta,a) \in [\underline{\theta},\theta'] \times [\underline{a},a']$ and 0 elsewhere. This function cannot be approximated by functions in $\mathcal{L}_+$. Therefore, even with sufficient variation in Z, $\mathcal{L}_+$ is not rich enough to provide sufficient data to show that $F(\cdot,\cdot)$ is point identified.

Careful observation of why $\mathcal{L}_+$ is not dense in $C_0(\cdot)$ suggests that the problem could be that class of sets generated by level curves in s does not generate rectangles -the building blocks of Borel $\sigma-$ algebra in $\mathbb{R}^n$. This further suggests yet another intuitive reason why the identification fails. As mentioned earlier, we are able to assign probability measure to all sets of the form $\{(\theta,a \in \mathcal{S}_{\theta a} : \mathbf{v}(\theta,a;z) \leq \tilde{v}\}$ for all $\tilde{v} \in \mathcal{S}_{V|Z}$, and therefore on the $\sigma-$algebra of the sets generated by these sets $\mathcal{B} = \sigma(\mathbf{v}^{-1}(\tilde{v}))$. Now, the question is the following: Can we then uniquely extend the measure defined on $\mathcal{B}$ to the entire Borel $\sigma-$ algebra? From the classical uniqueness and extension theory of a probability measure, if $\mathcal{B}$ is a $\pi-$system then it is sufficient to extend the measure uniquely. Note that $\pi-$ system is class of sets closed intersection. Since $\mathcal{B}$ is not a $\pi-$ system the sufficient condition fails. Unlike the result which uses mixture, this argument is only suggestive because it is only a sufficient condition for identification and is only intended to complement the previous arguments. Our nonidentification result is formalized by the following proposition:

Proposition 3: *Suppose a continuum of insurance coverages is offered to each insuree and all claims are observed. Under Assumption 1' and Assumption 2', $F(\cdot,\cdot)$ is not identified.*

4 Conclusion

In this paper, nonparametric identification of an insurance model with bi-dimensional private information is investigated. It is shown that if this risk is defined as a probability of an accident, the model is cannot be nonparametrically identified. When risk is defined as the probability of an accident, the variation in the claims data cannot be used because for any coverage, filing one claim is the same as filing ten claims, say, as far as identification is concerned. Aryal, Perrigne and Vuong (2009) define risk as the expected number of accident and the number of accidents is modeled as a poisson process, and use the variation in the claims data, once a coverage is chosen to identify the model. In most of the models of insurance, risk is defined as the probability of accident and that implicitly ignores the fact that an insuree can have multiple accidents. An insurance contract is written for a period of at least six months, and up to one year, and in that period an insuree could have more than one accidents. For an insurer, it is the expected number of accidents that is of interest and not just the probability of one accident.

It is also interesting to note that the identification approach adopted here, in particular the Lemma 1, is reminiscent of identification that relies on completeness assumption such as in Tallis and Chesson (1982), Newey and Powell (2003) and Hu and Schennach (2008), among others. In view of this result, one could pursue identification under some parametric restrictions on $F(\cdot,\cdot)$ and or could characterize partial identification of $F(\cdot,\cdot)$, both of which although important are not pursued in this paper.

References

[1] **Aryal, G. and I. Perrigne (2010)**: "An Insurance Model with Bi-Dimensional Screening," Working Paper, The Pennsylvania State University.

[2] **Aryal, G., I. Perrigne and Q. Vuong (2009)**: "Identification of Insurance Models with Multidimensional Screening," Working Paper, The Pennsylvania State Univeristy.

[3] **Blum, J. R and V. Susarla (1977)**: "Estimation of a mixing distribution function," *The Annals of Probability*, 5, 200-209.

[4] **Chiappori, P. A. and B. Salanie (2001)**: "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56-78.

[5] **Conway, J. B. (1985)**: " A Course in Functional Analysis," *Springer-Verlag, New York*.

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

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

[8] **Hu, Y. and S. M. Schennach (2008)**: "Instrument Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195-216.

[9] **Newey, W. K. and J.L. Powell (2003)**: "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.

[10] **Rothschild, M. and J. Stiglitz (1976)**: "Equilibrium in Competitive Insurance Markets: An Essay on The Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90, 629-649.

[11] **Tallis, G. M. and P. Chesson (1982)**: "Identifiability of Mixtures," *Journal of Australian Math. Soc. Ser.*, A 32, 339-348.