Supplemental Material to ‘Econometrics of Insurance with Multidimensional Types’

Gaurab Aryal
Boston University

Isabelle Perrigne
Rice University

Quang Vuong
New York University

Haiqing Xu
University of Texas Austin

September 2024

Correspondence to Isabelle Perrigne: iperrigne@gmail.com

The supplemental matertial contains three appendices. Appendix A presents the identification results when the damage distribution is truncated at the deductible. Appendix B considers alternative specifications to the CARA function and the Poisson distribution for the insurees' utility function and the number of accidents. Appendix C establishes several lemmas mentioned in the text or used in the appendices. Some can be of independent interest whereas others are likely known though we have yet found references for them.

Appendix A

This appendix extends the results of Section 3 when not all damages/accidents are observed because of truncation at the deductible. Not observing all the accidents limits the extent of identification. In particular, we show that $F(\cdot,\cdot|X)$ is identified up to the knowledge of the probability of damage below the lowest deductible, i.e., $H(dd_2(X)|X)$. Thus, when one of the insurance contracts offers full coverage, namely $dd_2(X)=0$, this probability is known, and the identification results of Section 3 apply despite truncation for insurees who choose $(t_1(X), dd_1(X))$. To simplify the notation, we let $H_c(X) \equiv H(dd_c(X)|X)$, $c=1,2$. Hereafter, we assume that $0 < dd_2(X) < dd_1(X) < \overline{d}(X)$ so that $0 < H_2(X) < H_1(X) < 1$.

We first derive a relation between $1-H_1(X)$ and $1-H_2(X)$, which allows us to focus on identification in terms of $H_2(X)$. Because a claim is filed only when the damage is above the deductible, from the observed claims, we identify the damage distribution conditional on the damage being larger than the deductible, i.e., the truncated damage distributions $H_c^*(\cdot|X) \equiv [H(\cdot|X) - H_c(X)]/[1-H_c(X))]$ on $[dd_c(X), \overline{d}(X)]$ from the insurees buying the coverage $(t_c(X), dd_c(X))$ for $c=1,2$. Differentiating and taking the ratio show that

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

for all $D \geq dd_1(X)$, where $0 < \lambda(X) < 1$ and $h_c^*(\cdot|X), c=1,2$ denotes the truncated damage density conditional on X. In particular, the function $\lambda(\cdot)$ is identified from the claim data, while $H(\cdot|X)$ is identified on $[dd_2(X), \overline{d}(X)]$ up to the knowledge of $H_2(X)$. We proceed in steps as in Section 3.

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

Let $\tilde{\theta} \equiv (1-H_2(X))\theta$ which replaces θ. We modify our identification argument as J is now unobserved. To identify the marginal density $f_{\tilde{\theta}|X}(\cdot|\cdot)$, we exploit the observed number of reported accidents $J^* \leq J$. The moment-generating function of J^* given (χ, X), where $\chi \in \{1,2\}$ indicates the insuree's coverage choice, is

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

where the third equality uses the moment-generating function of J^* given (J, χ, X), which is a Binomial $\mathcal{B}(J, 1 - H_\chi(X))$ from A2-(iv), and the fifth equality follows from A2-(iii) and the moment-generating function of the Poisson distribution. Thus,

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

for $u \in (-1 + H_\chi(X), +\infty)$, where $H_\chi(X) < 1$. Hence the distribution of risk θ given (χ, X) is identified up to the knowledge of $H_\chi(X)$.

Since $\tilde{\theta} = (1 - H_2(X))\theta$, its moment-generating function given (χ, X) is

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

for all $u \in (-\lambda(x), +\infty)$ and $u \in (-1, +\infty)$, respectively, where $\lambda(x) > 0$. Thus, the moment-generating function of $\tilde{\theta}$ given X is the weighted average

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

for $u \in (-\lambda(x), +\infty)$. Thus $f_{\tilde{\theta}|X}(\cdot|\cdot)$ is identified as $\lambda(X)$ is identified while $\nu_1(X)$ and $\nu_2(X)$ are the proportions of insurees choosing coverages 1 and 2 that are observed in the data. Since $f_{\theta|X}(\theta|x) = (1 - H_2(x))f_{\tilde{\theta}|X}((1 - H_2(x))\theta|X)$, the conditional distribution $F_{\theta|X}(\cdot|x)$ is identified up to $H_2(x)$.

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

Let $\tilde{\theta} \equiv (1 - H_2(X))\theta$ which replaces θ. We modify our identification argument as J is now unobserved. To identify the marginal density $f_{\tilde{\theta}|X}(\cdot|\cdot)$, we exploit the observed number of reported accidents J^*. The moment-generating function of J^* given (χ, X), where $\chi \in \{1, 2\}$ indicates the insuree's coverage choice, is

$$
\begin{aligned}
M_{J^*|\chi,X}(t|c,x) &= \mathrm{E}[e^{J^*t}|\chi = c, X = x] \\
&= \mathrm{E}\{\mathrm{E}[e^{J^*t}|J, \chi, X]|\chi = c, X = x\} \\
&= \mathrm{E}\left\{[H_\chi(X) + (1 - H_\chi(X))e^t]^J|\chi = c, X = x\right\} \\
&= \mathrm{E}\left\{\mathrm{E}[e^{J\log[H_\chi(X)+(1-H_\chi(X))e^t]}|\theta, a, \chi, X]|\chi = c, X = x\right\} \\
&= \mathrm{E}\left[e^{\theta[H_\chi(X)+(1-H_\chi(X))e^t-1]}|\chi = c, X = x\right] \\
&= M_{\theta|\chi,X}[(1 - H_\chi(X))(e^t - 1)|c, x],
\end{aligned} \tag{S.3}
$$

where the third equality uses the moment-generating function of J^* given (J, χ, X), which is a Binomial $\mathcal{B}(J, 1 - H_\chi(X))$ from A2-(iv), and the fifth equality follows from A2-(iii) and the moment-generating function of the Poisson distribution. Thus,

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

for $u \in (-1 + H_\chi(X), +\infty)$, where $H_\chi(X) < 1$. Hence the distribution of risk θ given (χ, X) is identified up to the knowledge of $H_\chi(X)$.

Since $\tilde{\theta} = (1 - H_2(X))\theta$, its moment-generating function given (χ, X) is

$$
\begin{aligned}
M_{\tilde{\theta}|\chi,X}(u|c,x) &= M_{\theta|\chi,X}(u(1 - H_2(x))|c, x) \\
&= \begin{cases} M_{J^*|\chi,X}\left[\log\left(1 + \frac{u}{\lambda(x)}\right)|1, x\right] & \text{if } c = 1, \\ M_{J^*|\chi,X}\left[\log\left(1 + u\right)|2, x\right] & \text{if } c = 2, \end{cases}
\end{aligned} \tag{S.4}
$$

for all $u \in (-\lambda(x), +\infty)$ and $u \in (-1, +\infty)$, respectively, where $\lambda(x) > 0$. Thus, the moment generating function of $\tilde{\theta}$ given X is the weighted average

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

for $u \in (-\lambda(x), +\infty)$. Thus $f_{\tilde{\theta}|X}(\cdot|\cdot)$ is identified as $\lambda(X)$ is identified while $\nu_1(X)$ and $\nu_2(X)$ are the proportions of insurees choosing coverages 1 and 2 that are observed in the data. Since $f_{\theta|X}(\theta|x) = (1 - H_2(x))f_{\tilde{\theta}|X}((1 - H_2(x))\theta|X)$, the conditional distribution $F_{\theta|X}(\cdot|x)$ is identified up to $H_2(x)$.

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

Following Section 3, we consider the probability that a (θ, a)-insuree with characteristics X chooses the coverage $(t_1(X), dd_1(X))$. Using (5) and $1 - H(D|X) = (1 - H_2(X))(1 - H_2^*(D|X))$, we remark that the indifference frontier between the two coverages in the space $(\tilde{\theta}, a)$ is given by

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

leading to the inverse $a(\tilde{\theta}, X)$, which is identified. From Bayes' rule, as in (6), we have

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

where $\nu_1(x)$ is observed and $f_{\tilde{\theta}|X}(\tilde{\theta}|x)$ is identified from the first step. Moreover, $f_{\tilde{\theta}|\chi,X}(\cdot|1, x)$ is identified because its moment-generating function $M_{\tilde{\theta}|\chi,X}(\cdot|1, x)$ is identified on $(-\lambda(x,),$ $+\infty)$ as shown in (S.2). Thus $F_{a|\tilde{\theta},X}(\cdot|\tilde{\theta}, x)$ is identified on the frontier $a(\tilde{\theta}, x)$.

Lastly, we note that $F_{a|\tilde{\theta},X}(a(\tilde{\theta}, x)|\tilde{\theta}, x) = F_{a|\theta,X}(a(\theta, x)|\theta, x)$ thereby identifying the latter up to $H_2(x)$ since $\tilde{\theta} = (1 - H_2(x))\theta$. Hence under the exclusion restriction and full support condition in A3, $F_{a|\theta,X}(\cdot|\cdot, \cdot)$ and hence $F(\theta, a|X, Z)$ are identified up to the knowledge of $H_2(X)$. This result is formally stated next.

Proposition 2: *Suppose that there are two offered coverages with* $0 < dd_2(X) < dd_1(X) < \overline{d}(X)$*, and damages are observed only when they are above the deductible for each insuree. Under A2–A3, the structure* $[F(\cdot, \cdot|X), H(\cdot|X)]$ *is identified up to* $H_2(X)$*.*

In the absence of the full support condition A3-(ii), the comments after Proposition 1 still apply, up to the knowledge of $H_2(X)$. We note that having a larger number of coverages $C > 2$ can only improve the identification results for two reasons. First, as in Section 3, more frontiers of the form (5) are more likely to cover the whole support $\Theta(x) \times \mathcal{A}(x_0)$ when Z varies. Second, the lowest deductible $dd_C(x)$ among the C coverages might not be binding as the probability of damage below $dd_C(x)$ is small. In practice, this can be assessed by estimating the conditional damage distribution in the neighborhood $[dd_C(X), (1+\delta)dd_C(X)]$ with δ small.

A consequence of Proposition 2 is that the structure $[F(\cdot, \cdot|X), H(\cdot|X)]$ is identified if and only if $H_2(X)$ is identified. The next result shows that $H_2(X)$ is not identified.

Proposition 3: *Under the conditions of Proposition 2,* $H_2(X)$ *is not identified.*

The proof relies on exhibiting an observationally equivalent structure, as shown below.

Proof of Proposition 3: Given Proposition 2, $H_2(X)$ is identified if and only if the structure $[F(\cdot,\cdot|X), H(\cdot|X)]$ is. Thus, it suffices to show that the latter is not identified. Let $[F(\cdot,\cdot|X), H(\cdot|X)]$ be a structure satisfying A2–A3. We construct a second structure $[\tilde{F}(\cdot,\cdot|X),\ \tilde{H}(\cdot|X)]$ as follows. Let $\kappa(\cdot)$ be a (measurable) function of X that can be arbitrarily large as long as $\kappa(x) \geq 1 - H_2(x)$ for all $x \in \mathcal{S}_X$. Let $\tilde{\theta} = \kappa(X)\theta$ and $\tilde{a} = a$. Since $\kappa(X) > 0$ from $1 - H_2(X) > 0$, we have

$$\tilde{f}(\cdot,\cdot|X) = \frac{1}{\kappa(X)} f\left(\frac{\cdot}{\kappa(X)}, \cdot \mid X\right).$$

Let $\tilde{h}(D|X) = [1/\kappa(X)]h(D|X)$ for $D \geq dd_2(X)$. Thus, $\int_{dd_2(X)}^{\overline{d}(X)} \tilde{h}(D|X)dD = [1-H_2(X)]/\kappa(X) \in (0,1]$. For $D \in [0, dd_2(X))$, define $\tilde{h}(D|X)$ to be nonnegative as long as $\int_0^{dd_2(X)} \tilde{h}(D|X)dD = 1 - [1 - H_2(X)]/\kappa(X)$. In particular, upon evaluating $\tilde{H}(D|X)$, it is easy to verify that

$$1 - \tilde{H}(D|X) = \frac{1}{\kappa(X)}[1 - H(D|X)] \tag{S.5}$$

for $D \geq dd_2(X)$. Moreover, the second structure $[\tilde{F}(\cdot,\cdot|X), \tilde{H}(\cdot|X)]$ satisfies A2–A3.

We now show that these two structures are observationally equivalent, i.e., they lead to the same distribution for the observables $(J^*, D_1^*, \ldots, D_{J^*}^*, \chi)$ given X and (t_1, dd_1, t_2, dd_2), where J^* and D^* refer to the number of reported accidents and their corresponding damages, respectively, while χ indicates the coverage chosen by the insuree. Regarding the distribution of $\tilde{\chi}$ given X, we note that $\tilde{\chi} = \chi$. The latter follows from $\tilde{\chi} = 1$ if and only if $(\tilde{\theta}, a) \in \tilde{\mathcal{C}}_1(X)$, i.e., $\tilde{\theta} \leq \tilde{\theta}(a, X)$. But $\tilde{\theta} = \kappa\theta$ while the frontier (5) defining the insurees' coverage choice for the second structure satisfies

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

using (S.3). Hence $\tilde{\theta} \leq \tilde{\theta}(a, X)$ is equivalent to $\theta \leq \theta(a, X)$. That is, $\tilde{\chi} = 1$ if and only if $\chi = 1$. Thus, the distributions of $\tilde{\chi}$ and χ given X are the same, i.e., $\tilde{\nu}_c(X) = \nu_c(X)$ for $c = 1, 2$.

Regarding the distribution of $\tilde{J}^*$ given $(\tilde{\chi},X)=(\chi,X)$, from (S.1), its moment-generating function is

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

where the first equality uses $M_{\tilde{\theta}|\chi,X}(u|c,x) = M_{\theta|\chi,X}[\kappa(x)u|c,x]$, which follows from $\tilde{\theta} = \kappa(X)\theta$, and $1-\tilde{H}_c(X) = (1-H_c(X))/\kappa(X)$, which follows from (S.3). Hence, the distribution of $\tilde{J}^*$ given (χ, X) is the same as that of J^* given (χ, X).

Lastly, regarding the distribution of reported damage $\tilde{D}^*$ given $(\tilde{J}^*, \chi, X)$ we have

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

on $[dd_2(X), \overline{d}(X)]$ where we have used $1 - \tilde{H}_\chi(\cdot|X) = (1 - H_\chi(\cdot|X))/\kappa(X)$ and (S.3). Hence, the two structures lead to the same distributions for the observables as desired. □

Proposition 3 shows that all the information provided by the model and the data have been exhausted. The nonidentification can be explained as follows. It arises from a compensation between the increase (decrease) in the number of accidents and an appropriate decrease (increase) in the probability of damages greater than the deductible. From the insuree's perspective, such compensation maintains the relative ranking between the two contracts. Thus, if a (θ, a)-insuree buys $(t_1(X), dd_1(X))$, then the $((1 - H_2(X))\theta, a)$-insuree also buys the same coverage if there is an appropriate increase in the probability of damages being greater than $dd_1(X)$. From the insurer's perspective, the decrease in the average number of accidents is compensated by an appropriate increase in the probability that the damage is above the deductible so that the expected payment to the insuree remains the same under either coverage.

Identification Strategies for $H_2(X)$

We discuss identification strategies for the probability $H_2(X)$. Any strategy that identifies $H_2(X)$ identifies the structure $[F(\cdot,\cdot|X),\ H(\cdot|X)]$. A first strategy is to parameterize the damage distribution $H(\cdot|X)$ as $H(\cdot|X;\beta)$ on $[0, \overline{d}(X)]$ with $\beta \in \mathcal{B} \subset I\!R^q$. Observations on reported damages D^* identify β and hence $H(\cdot|X)$ on $[0, \overline{d}(X)]$. Thus $H_2(X) \equiv H(dd_2(X)|X;\beta)$ is identified. In particular, we can choose a parametrization to fit the estimated truncated damage distribution $H^*(\cdot|X)$ on $[dd_2(X), \overline{d}(X)]$.

A second strategy is to consider additional data on the average number of accidents. For instance, suppose that for every $x \in \mathcal{S}_X$, we know the average number of accidents $\mu(x) \equiv \mathrm{E}[J|X = x] = \mathrm{E}\{\mathrm{E}[J|\theta, X = x]|X = x\} = \mathrm{E}[\theta|X = x]$ by A2-(iii). For the average number of reported accidents, we have $\mu^*_c(x) \equiv \mathrm{E}[J^*|\chi = c, X = x] = \mathrm{E}\{\mathrm{E}[J^*|J, \chi = c, X = x]|\chi = c, X = x\} = \mathrm{E}[J(1 - H_c(X))|\chi = c, X = x] = [1 - H_c(x)]\mathrm{E}[\theta|\chi = c, X = x]$ for $c = 1, 2$

since J^* given (J, χ, X) is distributed as a Binomial with parameters $(J, 1 - H_\chi(X))$. Thus

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

identifying $H_2(x)$, given that $\nu_c(x)$, $\mu_c^*(x), c = 1, 2$, and $\lambda(x)$ are identified from the data. Alternatively, suppose we know only $\mathrm{E}[J|X = x_*]$ for some x_*. Using the same argument establishes the identification of $H_2(x_*)$. This result combined with a support assumption such as $\overline{\theta}(x) = \overline{\theta}$ for every x identifies $H_2(x)$. To see this, note that $\overline{\tilde{\theta}}(x) = (1 - H_2(x))\overline{\theta}(x)$, where $\overline{\tilde{\theta}}(x)$ is the upper boundary of the support of $f_{\tilde{\theta}|X}(\cdot|X = x)$, which is identified. Applying this equation at x_* identifies $\overline{\theta} = \overline{\theta}(x_*) = \overline{\tilde{\theta}}(x_*)/(1 - H_2(x_*))$. Applying this equation again at different values x identifies $H_2(x)$. A similar argument applies at the lower bound $\underline{\theta}(x) = \underline{\theta}$.[1]

A third strategy is to derive sharp bounds on the probability $H_2(X)$. This approach, also known as partial identification, was developed by e.g. Manski and Tamer (2002), Haile and Tamer (2003) and Chernozhukov, Hong and Tamer (2007). Our bounds are nonparametric. Let $[F(\cdot, \cdot|X), H(\cdot|X)]$ be the structure generating the observables. Fix a value $x \in \mathcal{S}_X$. Proposition 2 implies that it is sufficient to determine the identified set for $H_2(x)$, i.e., the set of values $\tilde{H}_2(x)$ that are observationally equivalent to $H_2(x)$. To be precise, this is the set of values $\tilde{H}_2(x)$ corresponding to structures $[\tilde{F}(\cdot, \cdot|X), \tilde{H}(\cdot|X)]$ that are observationally equivalent to $[F(\cdot, \cdot|X), H(\cdot|X)]$. Indeed, Proposition 2 shows that $[F(\cdot, \cdot|X), H(\cdot|X)]$ is identified up to $H_2(\cdot)$. The proof of Proposition 3 above shows that any value $\tilde{H}_2(X) = 1 - (1/\kappa)[1 - H_2(x)]$ for $\kappa \in (1 - H_2(X), \infty)$ is observationally equivalent to $H_2(x)$. Thus, the identified set for $H_2(x)$ is $(0, 1)$.[2]

To tighten the upper bound, we can use some empirical evidence. From Cohen and Einav (2007), the estimated damage density decreases when the damage approaches the deductible from above, suggesting that the density below the deductible is not greater than its value at

[1]In contrast, information on the average damage is not sufficient. We note that $\mathrm{E}(D|X = x) = H_2(x)\mathrm{E}[D|D \leq dd_2(x), X = x] + (1 - H_2(x))\mathrm{E}[D|D \geq dd_2(x), X = x]$, where $\mathrm{E}[D|D \geq dd_2(x), X = x]$ is identified from the data. Thus, identification of $H_2(x)$ requires to know both $\mathrm{E}[D|D \leq dd_2(x), X = x]$ and $\mathrm{E}(D|X = x)$.

[2]We can then obtain sharp bounds for the structure $[F(\cdot, \cdot|\cdot), H(\cdot|\cdot)]$. Fixing $X = x$, let $H_2(x) = h \in (0, 1)$ leading to a unique structure $[F_h(\cdot, \cdot|x), H_h(\cdot|x)]$ by Proposition 2. The identified set for $[F(\cdot, \cdot|x), H(\cdot|x)]$ is the collection of such $[F_h(\cdot, \cdot|x), H_h(\cdot|x)]$ when h runs over $(0, 1)$. To derive bounds, we can follow Haile and Tamer (2003) by taking lower and upper envelopes for each structural distribution $F(\cdot, \cdot|x)$ or $H(\cdot|x)$.

the deductible. Thus we can assume that the damage density satisfies $h(D|x) \leq h[dd_2(x)|x]$ for every $D \leq dd_2(x)$ and x. Integrating both sides from 0 to $dd_2(x)$ we obtain $0 \leq H_2(x) \leq dd_2(x)h(dd_2(x)|x)$. Dividing both sides by $1-H_2(x)$, and using the definition of the truncated density $h_2^*(\cdot|x)$, we obtain

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

Solving for $H_2(x)$ gives the upper bound

$$H_2(x) \leq \frac{dd_2(x)h_2^*(dd_2(x)|x)}{1+dd_2(x)h_2^*(dd_2(x)|x)} \equiv \overline{H}_2(x),$$

which is strictly less than 1. Thus the identified set for $H_2(x)$ reduces to $(0, \overline{H}_2(x)]$. The upper bound can be estimated from observables.[3]

The extension of Section 3 applies up to the identification of $H_2(\dot{})$. It is worth noting that even if damages were not observed below the deductibles, condition (8) would still be implementable and verifiable. Indeed, upon dividing by $[1 - H(dd_C)]$ the ratio in the RHS is equal to $\int_{dd_{c+2}}^{dd_{c+1}}[1 - H_C^*(D)]dD / \int_{dd_{c+1}}^{dd_c}[1 - H_C^*(D)]dD$ where $H_C^*(\cdot) = [H(\cdot) - H(dd_C)]/[1 - H(dd_C)]$ is the distribution of D conditional on $D > dd_C$. The latter distribution is identified from the claims of individuals buying the highest coverage (t_C, dd_C), i.e., the lowest deductible dd_C. Thus the Corollary still applies, and one should observe that the observed coverages should lie on a convex curve in the (t, dd)-space. The estimation procedure of Section 4 extends to this case with some appropriate adjustments, and the analyst can choose an estimator for $H_2(\cdot)$ in line with the identification strategies discussed above.

Appendix B

This appendix extends the results of Section 3 under alternative specifications to the CARA function and the Poisson distribution for the insurees' utility function and the number of accidents. As before, we assume that heterogeneity across insurees is aracterized by a bidimensional vector (θ, a) of private information. Thus, by necessity parametric assumptions on the utility function and the distribution of the number of accidents follow. Specifically, we let the increasing and concave utility function $U(\cdot; a)$ be parameterized by $a \in I\!R_+$ capturing

[3]The comment in the previous footnote applies with h running over $(0, \overline{H}_2(x)]$.

the insuree's risk aversion, while the distribution of the number of accidents $P(\cdot;\theta)$ is parameterized by θ capturing the insuree's risk. These need no longer be the CARA utility function nor the Poisson distribution. We stress that the CARA-Poisson specification is widely used for its mathematical tractability that we loose under alternative specifications. We still leave the joint distribution of (θ, a) unspecified. Thus our approach remains semiparametric as we consider nonparametric mixtures for the utility function and the distribution of the number of accidents in the population.

We extend A1 by replacing A1-(i,iii) with more general assumptions on the utility function and the distribution of the number J of accidents.

Assumption A1': *While A1-(ii,iv) remain, A1-(i,iii) are replaced by*
(i) The insuree's utility function $U(x;a)$ is increasing and concave in x, where $U(\cdot;a')$ is more risk averse than $U(\cdot;a)$ whenever $a' > a$,[4]
(iii) The number of accidents J given θ is distributed as $P(\cdot;\theta)$, where $P(\cdot;\theta')$ First-Order Stochastically Dominates (FOSD) $P(\cdot;\theta)$, i.e., $P(\cdot;\theta') \overset{FOSD}{\succ} P(\cdot;\theta)$, whenever $\theta' > \theta$.

The main role of A1'-(i,iii) is to rank insurees in terms of their risk aversion and risk parameters a and θ, respectively. Beside the CARA utility function, a well-known utility function satisfying A1'-(i) is the Constant Relative Risk Aversion (CRRA) utility function $x^{1-a}/[1-a]$ for $a \in (0,1)$. More generally, we can consider the Hyperbolic Absolute Risk Aversion (HARA) utility function which nests ARA and RRA as special cases. It is defined as

$$U(x;\alpha,\beta,\gamma) = \frac{1-\alpha}{\alpha}\left(\frac{\gamma x}{1-\alpha} + \beta\right)^{\alpha},$$

where $\alpha \neq 0$, $\gamma > 0$ and $[\gamma x/(1-\alpha)] + \beta > 0$. In general, we expect $\alpha < 1$ since $\alpha > 1$ would correspond to implausible Increasing ARA (IARA) while $\beta \in I\!R$. The family of HARA utility functions has three parameters (α, β, γ), whereas heterogeneity in risk aversion is one-dimensional in our setting. We can achieve the latter by fixing two of these three utility parameters, thereby leaving the third one to describe individuals' heterogeneity in risk aversion. For instance, the CARA specification $U(x;a) = -\exp(-ax)$ corresponds to $\beta = [1-(1/\alpha)]^{-1/\alpha}$ and $\gamma = a[1-(1/\alpha)]^{1-(1/\alpha)}$ with $\alpha \to \pm\infty$. See also Lemma C.4 in Appendix C. Alternatively, the CRRA specification $U(x;a) = x^{1-a}/(1-a)$ corresponds to $\beta = 0$ and $\gamma = (1-\alpha)^{1-(1/\alpha)}$. A strictly negative value of β implies Decreasing RRA (DRRA)

[4]By definition, this means that there exists an increasing and concave function $q(\cdot)$ possibly depending on (a', a) such that $U(\cdot;a') = q[U(\cdot;a)]$.

preferences, while a strictly positive value of β implies Increasing RRA (IRRA) preferences. These specifications satisfy A1'-(i) and allow the nature of risk aversion a to vary across individuals. Regarding A1'-(iii), several families of distributions satisfy it such as the Poisson distribution $\mathcal{P}(\cdot;\theta)$ for $\theta > 0$, the Negative Binomial distribution $\mathcal{NB}(\theta, p)$ for $\theta > 0$ and fixed $p \in (0, 1)$, and the Binomial distribution $\mathcal{B}(n, \theta)$ for $\theta \in (0, 1)$ and fixed $n \geq 1$. See Lemma C.2 in Appendix C.

We can now define the certainty equivalent $CE(t, dd; \theta, a, w)$ for a (θ, a)-individual with wealth w. Following (1) and (2), $CE(t, dd; \theta, a, w)$ solves

$$U(CE; a) \equiv \mathrm{E}\left[U\Big(w - t - \sum_{j=0}^{J} \min\{dd, D_j\}; a\Big)|\theta\right], \tag{S.6}$$

where $J \sim P(\cdot;\theta)$, $D_j \overset{iid}{\sim} H(\cdot)$ and $D_0 \equiv 0$ by convention. Thus the frontier $\theta_{c,c+1}(\cdot)$ between coverages (t_c, dd_c) and (t_{c+1}, dd_{c+1}) satisfies

$$CE(t_c, dd_c; \theta_{c,c+1}(a), a, w) - CE(t_{c+1}, dd_{c+1}; \theta_{c,c+1}(a), a, w) = 0 \tag{S.7}$$

for all $a \in [\underline{a}, \overline{a}]$. The frontier $\theta_{c,c+1}(\cdot)$ actually depends on wealth w. To simplify the notation, we omit this dependence. We also let $CE_c(\theta, a) \equiv CE(t_c, dd_c; \theta, a, w)$.

The next lemma extends Lemma 1 while allowing for more than two contracts.

Lemma 1': *Let A1' hold. Let the coverages (t_c, dd_c), $c = 1, \ldots, C \geq 2$ satisfy the RP condition (7).*

(i) When $dd_c = 0$ (full coverage), $CE_c(\theta, a)$ reduces to $w - t_c$. When $dd_c > 0$, $CE_c(\theta, a)$ decreases in both risk θ and risk aversion a.

(ii) Suppose in addition that $CE_c(\theta, a)$ is supermodular in (c, θ) and (c, a).[5] *Conditional on wealth w, the frontier $\theta_{c,c+1}(\cdot)$ between the coverages (t_c, dd_c) and (t_{c+1}, dd_{c+1}) is decreasing in the (θ, a)-space. Every (θ, a)-individual below (resp. above) this frontier prefers coverage c over coverage $c+1$ (resp. $c+1$ over c).*

Proof: (i) When $dd = 0$, (S.6) reduces to $U(CE; a) = \mathrm{E}[U(w - t; a)] = U[w - t; a)]$. Thus $CE(t, dd; \theta, a, w) = w - t$ since $U(\cdot; a)$ is increasing by A1'-(i). Next, consider the

[5]A function $\psi(c, x, y)$ is supermodular in (c, x) whenever $\psi(c', x', y) + \psi(c, x, y) > \psi(c', x, y) + \psi(c, x', y)$ for all $c' > c$, $x' > x$ and y. If $\psi(c, x, y)$ is differentiable in x this is equivalent to $\partial\psi(c, x, y)/\partial x$ increasing in c for all (c, x, y), or $\partial\psi(c', x, y)/\partial x - \partial\psi(c, x, y)/\partial x > 0$ for all $c' > c$ and (x, y). If $\psi(c, x, y)$ was differentiable in (c, x), this would be also equivalent to $\partial^2\psi(c, x, y)/\partial c\partial x > 0$. See Topkis (1978).

case $dd > 0$. Let $J' \sim P(\cdot;\theta')$ and $J \sim P(\cdot;\theta)$ where $\theta' > \theta$ so that $P(\cdot;\theta') \overset{FOSD}{\succ} P(\cdot;\theta)$ by A1'-(iii). Setting $X_j \equiv \min\{dd, D_j\}$ for $j \in \{0,1,2,\ldots\}$ in Lemma C.6, we have $\sum_{j=0}^{J'} \min\{dd, D_j\} \overset{FOSD}{\succ} \sum_{j=0}^{J} \min\{dd, D_j\}$. Hence $w - t - \sum_{j=0}^{J} \min\{dd, D_j\} \overset{FOSD}{\succ} w - t - \sum_{j=0}^{J'} \min\{dd, D_j\}$. Because $U(\cdot;a)$ is increasing by A1'-(i), it follows from e.g. Gollier (2001) that $\mathrm{E}\left[U(w - t - \sum_{j=0}^{J} \min\{dd, D_j\}; a)\right] > \mathrm{E}\left[U(w - t - \sum_{j=0}^{J'} \min\{dd, D_j\}; a)\right]$. That is, the certainty equivalent $CE(t, dd;\ \theta, a, w)$ defined by (S.6) is decreasing in θ. It remains to verify that $CE(t, dd; \theta, a, w)$ is decreasing in a. Indeed because the a'-individual is more risk averse than the a-individual when $a' > a$ by A1'-(i), it follows that the former is worse off than the latter when offered the same lottery $w - t - \sum_{j=0}^{J} \min\{dd, D_j\}$. See Rothschild and Stiglitz (1970) or e.g. Gollier (2001).

(ii) The frontier between the two coverages (t_c, dd_c) and (t_{c+1}, dd_{c+1}) is the locus of (θ, a) pairs satisfying the indifference condition (S.7), i.e.,

$$CE_c(\theta, a) - CE_{c+1}(\theta, a) = 0. \tag{S.8}$$

Let (θ, a) and (θ', a') be two points on this locus with $a' > a$. We want to show that $\theta' < \theta$. Because $CE_c(\theta, a)$ is supermodular in (c, a), we have $CE_c(\theta, a') - CE_c(\theta, a) < CE_{c+1}(\theta, a') - CE_{c+1}(\theta, a)$ or upon rearranging terms

$$CE_c(\theta, a') - CE_{c+1}(\theta, a') < CE_c(\theta, a) - CE_{c+1}(\theta, a) = 0 \tag{S.9}$$

using (S.8). In particular, (S.9) shows that $\theta' \neq \theta$ since (θ', a') being on the indifference locus satisfies $CE_c(\theta', a') - CE_{c+1}(\theta', a') = 0$. Moreover, subtracting the latter equation from (S.9) gives upon rearranging terms

$$CE_{c+1}(\theta', a') - CE_{c+1}(\theta, a') < CE_c(\theta', a') - CE_c(\theta, a').$$

Because $CE_c(\theta, a)$ is supermodular in (c, θ), we would obtain the reverse inequality if $\theta' > \theta$. Thus $\theta' \leq \theta$ and hence $\theta' < \theta$ since $\theta' \neq \theta$ as shown earlier.

It remains to show the second part of (ii). It suffices to show that any individual above or equivalently to the right of the indifference locus prefers coverage $c + 1$. Let (θ', a') be such an individual and (θ', a) be the corresponding pair on the indifference locus with $a < a'$. Similarly to (S.9) with θ replaced by θ' we obtain $CE_c(\theta', a') - CE_{c+1}(\theta', a') < 0$. Thus the (θ', a')-individual prefers coverage 2 as desired.[6] □

[6] An alternative proof uses the differentiability of $CE_c(\theta, a)$ in (θ, a) for $c = 1, 2$. Specifically, the

Part (i) of Lemma 1' conforms with intuition. If an individual has a larger risk θ, he/she will likely be involved in more accidents that will reduce his/her net wealth and thus his/her certainty equivalent. Likewise, an individual with a higher risk aversion a will have a lower utility and hence lower certainty equivalent for the coverage (t, dd) *ceteris paribus*. Regarding (ii), supermodularity captures the idea that a better coverage is more valuable to insurees with higher θ (resp. higher a) than to insurees with lower θ (resp. lower a). Supermodularity conditions are widely used in economic theory. See Milgrom and Roberts (1990), Vives (1990), and Athey (2002) among others. In our case, this condition involves the terms of the coverages, the utility function, the distribution of the number of accidents, and the distribution of damages since the expectation in (S.6) is taken with respect to $(J, D_1, \ldots, D_J)$.[7]

The next lemma extends Lemma 2 by relaxing the CARA-Poisson assumptions. As in Lemma 2, it ensures that the C frontiers $\theta_{c,c+1}(\cdot)$ do not cross and lie on top of each other as c increases from 1 to $C-1$.

Lemma 2': *Let A1' hold. Let the coverages (t_c, dd_c), $c = 1, \ldots, C \geq 2$ satisfy the RP condition (7), and $CE_c(\theta, a)$ be supermodular in (c, θ) and (c, a). Suppose that*

$$\frac{\frac{\partial CE_c(\theta_{c+1,c+2}(a),a)}{\partial a} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial a}}{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial a} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial a}} > \frac{\frac{\partial CE_c(\theta_{c+1,c+2}(a),a)}{\partial \theta} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial \theta}}{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial \theta} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial \theta}} \quad \text{(S.10)}$$

for $a \in [\underline{a}, \overline{a}]$ and $c = 1, \ldots, C-2$. Then the frontiers $\theta_{c,c+1}(\cdot)$ between coverages (t_c, dd_c) and $(t_{c+1},\, d_{c+1})$ for $c = 1, \ldots, C-1$ satisfy $\theta_{1,2}(\cdot) < \ldots < \theta_{C-1,C}(\cdot)$ on $[\underline{a}, \overline{a}]$ if and only if

$$CE_c(\theta_{c+1,c+2}(\underline{a}), \underline{a}) < CE_{c+1}(\theta_{c+1,c+2}(\underline{a}), \underline{a}) \quad \text{(S.11)}$$

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

Proof: Fix $c = 1, \ldots, C-2$. We want to show that $\theta_{c,c+1}(\cdot) < \theta_{c+1,c+2}(\cdot)$ on $[\underline{a}, \overline{a}]$. From (S.7) $\theta_{c,c+1}(\cdot)$ satisfies

$$CE_c(\theta_{c,c+1}(a), a) - CE_{c+1}(\theta_{c,c+1}(a), a) = 0. \quad \text{(S.12)}$$

first part of (ii) can be proved by totally differentiating (A.3) to obtain the derivative $\theta'(a)$ of the frontier and using the differentiable version of the supermodularity of $CE_c(\theta, a)$ in (c, θ) and (c, a). To prove the second part of (ii), we can use $CE_c(\theta', a') - CE_c(\theta', a) = \int_a^{a'} \partial CE_c(\theta', \tilde{a})/\partial a \, d\tilde{a}$ where $\partial CE_c(\theta', \tilde{a})/\partial a$ is increasing in c.

[7]As a matter of fact, Lemma C.3 in Appendix C establishes the supermodularity of $CE_c(\theta, a)$ in the CARA-Poisson specification of Section 2 irrespective of the damage distribution. When the utility function is HARA, the difficulty arises from the non-explicit form of the certainty equivalent even when the distribution of the number of accidents is Poisson. See Lemma C.4.

Because $CE_c(\theta, a)$ is supermodular in (c, θ), the derivative $\partial CE_c(\theta, a)/\partial\theta$ is increasing in c. Thus $\partial CE_c(\theta, a)/\partial\theta - \partial CE_{c+1}(\theta, a)/\partial\theta < 0$. Hence $CE_c(\cdot, a) - CE_{c+1}(\cdot, a)$ is decreasing. Thus, using (S.12), we have $\theta_{c,c+1}(a) < \theta_{c+1,c+2}(a)$ if and only if

$$CE_c(\theta_{c+1,c+2}(a), a) - CE_{c+1}(\theta_{c+1,c+2}(a), a) < 0. \tag{S.13}$$

We show that the LHS of (S.13) decreases in a under condition (S.10). Its (total) derivative with respect to a is

$$\begin{aligned} &\frac{d}{da}\Big[CE_c(\theta_{c+1,c+2}(a), a) - CE_{c+1}(\theta_{c+1,c+2}(a), a)\Big] \\ &\quad= \left[\frac{\partial CE_c(\theta_{c+1,c+2}(a), a)}{\partial\theta} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a), a)}{\partial\theta}\right]\theta'_{c+1,c+2}(a) \\ &\qquad+ \left[\frac{\partial CE_c(\theta_{c+1,c+2}(a), a)}{\partial a} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a), a)}{\partial a}\right], \end{aligned} \tag{S.14}$$

where $\theta'_{c+1,c+2}(\cdot)$ is the derivative of the frontier $\theta_{c+1,c+2}(\cdot)$. Because this frontier satisfies $CE_{c+1}(\theta_{c+1,c+2}(a), a) - CE_{c+2}(\theta_{c+1,c+2}(a), a) = 0$ by definition, differentiation with respect to a gives

$$\theta'_{c+1,c+2}(a) = -\frac{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial a} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial a}}{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial\theta} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial\theta}}.$$

Hence (S.14) gives upon rearranging terms

$$\begin{aligned} &\frac{d}{da}\Big[CE_c(\theta_{c+1,c+2}(a), a) - CE_{c+1}(\theta_{c+1,c+2}(a), a)\Big] \\ &= -\left[\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a), a)}{\partial a} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a), a)}{\partial a}\right] \\ &\times \left[\frac{\frac{\partial CE_c(\theta_{c+1,c+2}(a),a)}{\partial\theta} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial\theta}}{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial\theta} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial\theta}} - \frac{\frac{\partial CE_c(\theta_{c+1,c+2}(a),a)}{\partial a} - \frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial a}}{\frac{\partial CE_{c+1}(\theta_{c+1,c+2}(a),a)}{\partial a} - \frac{\partial CE_{c+2}(\theta_{c+1,c+2}(a),a)}{\partial a}}\right] \end{aligned}$$

The term within the first pair of brackets is negative since $CE_c(\theta, a)$ is supermodular in (c, a), so that $\partial CE_c(\theta, a)/\partial a$ is increasing in c. The term within the second pair of brackets is negative by condition (S.10). Thus the derivative of the LHS of (S.13) with respect to a is negative as desired. Because $CE_c(\theta_{c+1,c+2}(a), a) - CE_{c+1}(\theta_{c+1,c+2}(a), a)$ is decreasing in a, (S.13) holds for all $a \in [\underline{a}, \overline{a}]$ if and only if it holds at $\underline{a}$. □

Because $CE_c(\theta, a)$ is supermodular in (c, θ) and (c, a), numerators and denominators in (S.10) are negative. Note that (S.10) is evaluated at the frontier $\theta_{c+1,c+2}(\cdot)$. Thus, a sufficient condition for (S.10) is

$$\frac{\frac{\partial CE_c(\theta,a)}{\partial a} - \frac{\partial CE_{c+1}(\theta,a)}{\partial a}}{\frac{\partial CE_c(\theta,a)}{\partial\theta} - \frac{\partial CE_{c+1}(\theta,a)}{\partial\theta}} > \frac{\frac{\partial CE_{c+1}(\theta,a)}{\partial a} - \frac{\partial CE_{c+2}(\theta,a)}{\partial a}}{\frac{\partial CE_{c+1}(\theta,a)}{\partial\theta} - \frac{\partial CE_{c+2}(\theta,a)}{\partial\theta}} \tag{S.15}$$

for $(\theta, a) \in [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}]$ and $c = 1, \ldots, C-2$.[8] Moreover, it follows from (S.7) that the LHS and RHS of (S.15) are the slopes in absolute values of the frontiers $\theta_{c,c+1}(\cdot)$ and $\theta_{c+1,c+2}(\cdot)$, respectively, at (θ, a). If these frontiers intersect at (θ, a), (S.15) implies that they would intersect at most once. As a matter of fact, under (S.10), condition (S.11) ensures that the C frontiers $\theta_{c,c+1}(\cdot)$, $c = 1, \ldots, C-1$ do not intersect.

Condition (S.11) generalizes condition (8) of Lemma 2 to the case when the utility function and the distribution of the number of accidents are no longer CARA and Poisson, respectively. Indeed, using (2) to evaluate the difference $CE_c(\theta, a) - CE_{c+1}(\theta, a)$ at $(\theta_{c+1,c+2}(\underline{a}), \underline{a})$, where the frontier $\theta_{c+1,c+2}(\cdot)$ is given by (3), it is easy to verify that condition (S.11) reduces to condition (8). Moreover, from (S.6), condition (S.11) is equivalent to

$$\mathrm{E}\left[U\Big(w - t_c - \sum_{j=0}^{J_{c+1,c+2}(\underline{a})} \min\{dd_c, D_j\}; \underline{a}\Big)\right] < \mathrm{E}\left[U\Big(w - t_{c+1} - \sum_{j=0}^{J_{c+1,c+2}(\underline{a})} \min\{dd_{c+1}, D_j\}; \underline{a}\Big)\right]$$

for $c = 1, \ldots, C-2$, where $J_{c+1,c+2}(\underline{a}) \sim P(\cdot; \theta_{c+1,c+2}(\underline{a}))$. Since $U(\cdot; \underline{a})$ is increasing and concave, it follows that a sufficient condition for (S.11) is

$$-t_c - \sum_{j=0}^{J_{c+1,c+2}(\underline{a})} \min\{dd_c, D_j\} \overset{SOSD}{\prec} -t_{c+1} - \sum_{j=0}^{J_{c+1,c+2}(\underline{a})} \min\{dd_{c+1}, D_j\} \tag{S.16}$$

by definition of Second-Order Stochastic Dominance (SOSD). See Rothschild and Stiglitz (1970), or e.g., Gollier (2001). In particular, (S.16) is independent of wealth w. Furthermore, upon taking expectations and invoking A1-(iv), (S.16) requires that $t_{c+1} - t_c < \mathrm{E}[J_{c+1,c+2}(\underline{a})]\mathrm{E}\Big[\min\{dd_c, D\} - \min\{dd_{c+1}, D\}\Big]$, i.e.,

$$t_{c+1} - t_c < \mathrm{E}[J_{c+1,c+2}(\underline{a})] \int_{dd_{c+1}}^{dd_c} [1 - H(D)]dD, \tag{S.17}$$

where we have used the identity $\min\{dd_c, D\} = \min\{dd_{c+1}, D\} + [\min\{dd_c, D\} - dd_{c+1}]$ $1\!\!1(dd_{c+1} \le D)$ since $dd_c > dd_{c+1}$. Because the expectation $\mathrm{E}[J_{c+1,c+2}(\underline{a})]$ depends on the distribution of damage $H(\cdot)$, the lowest risk aversion $\underline{a}$ as well as the terms (t_{c+1}, dd_{c+1}) and (t_{c+2}, dd_{c+2}) of the coverages $c+1$ and $c+2$, (S.17) is related to reverse nonlinear pricing as for the CARA-Poisson case.[9]

[8]Condition (S.15) and hence (S.10) are satisfied by the CARA-Poisson specification of Section 2 irrespective of the damage distribution. See Lemma C.5 in Appendix C.

[9]In the latter case, $E[J_{c+1,c+2}(\underline{a})] = \theta_{c+1,c+2}(\underline{a})$. From (3), it follows that (S.17) reduces to condition (9) by letting $\underline{a}$ approaching zero.

Lastly, we show how to extend the identification results of Section 3. In line with A2, we make the following assumption.

Assumption A2': *While A2-(ii,iv) remain, A2-(i,iii) are replaced by A1'-(i,iii). In addition, the family of distributions $P(\cdot;\theta)$ is additively closed.*[10]

The distribution of accidents given X is a mixture of the distribution of the number J of accidents given θ with mixing distribution given by $F_{\theta|X}(\cdot|\cdot)$, i.e., $\Pr[J \leq \cdot|X] = \int P(\cdot|\theta) dF(\theta|X)$ using A2'-(iii). The additive closedness of $P(\cdot;\theta)$ then ensures the identification of the distribution $F_{\theta|X}(\cdot|\cdot)$ from the distribution of the number J of accidents given X. See Teicher (1961) and Rao (1992). Several families of discrete distributions are additively closed. For instance, the Poisson distribution $\mathcal{P}(\theta)$ for $\theta > 0$ and the Negative Binomial distribution $\mathcal{NB}(r,p)$ for $r > 0$ and fixed $p \in (0,1)$ mentioned earlier are additively closed. In contrast, the Binomial distribution $\mathcal{B}(n,p)$ for $p \in (0,1)$ and fixed $n \geq 1$ is not.[11] Given the identification of $F_{\theta|X}(\cdot|\cdot)$ under A2', we use steps 2 and 3 of Section 3.[12] Hence the joint distribution $F(\theta, a|X)$ of risk and risk aversion is identified under the exclusion and support assumption A3. As noted above, the latter can be weakened as it suffices that the combined variations of the $C-1$ frontiers span the $\Theta(X) \times \mathcal{A}(X_0)$ space.

Appendix C

This appendix establishes several lemmas mentioned in the text or used in Appendix B above. Some can be of independent interest. Others are likely known though we have yet found references for them.

Lemma C.1: *Let $H(\cdot)$ be a distribution function with support $(0,\overline{y})$. For any $y_\dagger \in (0,\overline{y})$, the function $\psi(x,y) \equiv \int_{y_\dagger}^{y} e^{xz}[1-H(z)]dz$ is log-supermodular in $(x,y) \in (0,+\infty) \times (y_\dagger, \overline{y})$.*

[10] A family of distributions $\{P(\cdot;\theta); \theta \in I\!R\}$ is additively closed if their characteristic functions $\phi(\cdot;\theta)$ satisfy $\phi(\cdot;\theta_1)\phi(\cdot;\theta_2) = \phi(\cdot;\theta_1+\theta_2)$ for every $(\theta_1,\theta_2) \in I\!R^2$. This is equivalent to $\phi(\cdot;\theta) = \phi(\cdot;1)^\theta$. See Rao (1992).

[11] When $n = 1$, i.e., when one only observes whether an individual has an accident or not, Aryal, Perrigne and Vuong (2012) show that our insurance model with CARA utility function is not identified despite exploiting all the restrictions of the model.

[12] The car value as a proxy for wealth w is included in X.

Proof of Lemma C.1: Following Topkis (1978), the function $\psi(x,y)$ is log-supermodular if $\partial^2 \log\psi(x,y)/\partial x\partial y > 0$. Taking derivatives, we obtain

$$\begin{aligned}\frac{\partial \log\psi(x,y)}{\partial y} &= \frac{e^{xy}[1-H(y)]}{\psi(x,y)}\\ \frac{\partial^2 \log\psi(x,y)}{\partial x\partial y} &= \frac{ye^{xy}[1-H(y)]\psi(x,y) - e^{xy}[1-H(y)]\int_{y_\dagger}^{y} ze^{xz}[1-H(z)]\,dz}{\psi(x,y)^2}.\end{aligned}$$

But $\int_{y_\dagger}^{y} ze^{xz}[1-H(z)]\,dz < y\int_{y_\dagger}^{y} e^{xz}[1-H(z)]\,dz = y\psi(x,y)$ since $z \le y$ and $y_\dagger < y$. Thus the numerator of $\partial^2 \log\psi(x,y)/\partial x\partial y$ is positive as desired. □

Lemma C.2: *The Poisson distribution $\mathcal{P}(\theta)$ for $\theta > 0$, the Negative Binomial distribution $\mathcal{NB}(r,p)$ for $r > 0$ and fixed $p \in (0,1)$, and the Binomial distribution $\mathcal{B}(n,p)$ for $p \in (0,1)$ and fixed $n \ge 1$ are families of increasingly FOSD distributions as θ, r and p increase respectively.*

Proof of Lemma C.2: From e.g., Johnson, Kemp and Kotz (2005), the Poisson cdf with parameter θ is given by $\Pr[X \le j] = 1 - \Pr[\text{Gamma}(j+1) \le \theta]$ which is decreasing in θ for every $j = 0,1,\ldots$. Thus $\mathcal{P}(\theta') \overset{FOSD}{\succ} \mathcal{P}(\theta)$ whenever $\theta' > \theta$. Similarly, the cdf of $\mathcal{NB}(r,p)$ is given by $\Pr[X \le j] = \Pr[\text{Beta}(r, j+1) \le p]$, which is decreasing in r for every $j = 0,1,\ldots$. To see the latter, we note that $\text{Beta}(a,b) \overset{d}{=} 1 - \text{Gamma}(b)/[\text{Gamma}(a) + \text{Gamma}(b)]$ where $\text{Gamma}(a)$ is independent of $\text{Gamma}(b)$. Since $\text{Gamma}(a') \overset{d}{=} \text{Gamma}(a) + \text{Gamma}(a'-a)$ where $\text{Gamma}(a)$ is independent of $\text{Gamma}(a'-a) \ge 0$, it follows that $\text{Gamma}(a') \overset{FOSD}{\succ} \text{Gamma}(a)$ for $a' > a$ implying that $\Pr[\text{Beta}(r', j+1) \le p] < \Pr[\text{Beta}(r, j+1) \le p]$ for $r' > r$. That is, $\mathcal{NB}(r',p) \overset{FOSD}{\succ} \mathcal{NB}(r,p)$ whenever $r' > r$. Lastly, the cdf of $\mathcal{B}(n,p)$ is given by $\Pr[X \le j] = \Pr[\text{Beta}(n-j, j+1) \le 1-p]$ which is decreasing in p for every $j = 0,1,\ldots,n$. Thus $\mathcal{B}(n,p') \overset{FOSD}{\succ} \mathcal{B}(n,p)$ whenever $p' > p$. □

Lemma C.3: *Let the coverages (t_c, dd_c), $c = 0,1,\ldots,C \ge 2$, satisfy the RP condition (7). The certainty equivalent $CE_c(\theta,a)$ of contract c is supermodular in (c,θ) and (c,a) when $U(\cdot;a)$ and $P(\cdot;\theta)$ are the CARA utility and the Poisson distribution, respectively.*

Proof of Lemma C.3: We want to show that

$$\frac{\partial CE_{c'}(\theta,a)}{\partial\theta} - \frac{\partial CE_c(\theta,a)}{\partial\theta} > 0 \quad \text{and} \quad \frac{\partial CE_{c'}(\theta,a)}{\partial a} - \frac{\partial CE_c(\theta,a)}{\partial a} > 0$$

for $c < c'$. To prove the first inequality, we note that (2) gives

$$\frac{\partial CE_{c'}(\theta,a)}{\partial\theta} - \frac{\partial CE_c(\theta,a)}{\partial\theta} = -\frac{1}{a}[\phi_a(dd') - \phi_a(dd)].$$

Because $\phi_a(\cdot)$ is increasing and $dd' < dd$, the desired inequality follows. Regarding the second inequality, we show that $\partial CE_c(\theta, a)/\partial a$ increases in c or equivalently decreases in dd. From (2) this is equivalent to showing that

$$\frac{\partial CE_c(\theta, a)}{\partial a} = -\theta \frac{\partial}{\partial a}\left[\frac{\phi_a(dd) - 1}{a}\right]$$

is decreasing in dd. But $[\phi_a(dd) - 1]/a = \int_0^{dd} e^{aD}[1 - H(D)]dD$ by (A.1). Thus,

$$\frac{\partial^2}{\partial dd \partial a}\left[\frac{\phi_a(dd) - 1}{a}\right] = \frac{\partial}{\partial a}\left[e^{add}[1 - H(dd)]\right] = ae^{add}[1 - H(dd)] > 0.$$

Hence $\partial CE_c(\theta, a)/\partial a$ is decreasing in dd as desired. □

Lemma C.4: *Let $U(x; a, \phi, \delta) = -\delta(1 - \phi ax)^{1/\phi}$ for $a > 0$, $\phi < 1$, $\delta > 0$ and $1 - \phi ax > 0$. Then the following holds*

(i) $U(\cdot; a, \phi, \delta)$ is a HARA utility. The CARA utility is obtained as $(\phi, \delta) \to (0, 1)$.

(ii) For any fixed $\phi < 1$ and $\delta > 0$, $U(x; a, \phi, \delta)$ satisfies Assumption 1'-(i).

Moreover, let Assumption A1'-(ii,iii,iv) holds with $P(\cdot; \theta)$ as the Poisson distribution and the coverages (t_c, dd_c), $c = 0, 1, \ldots, C \geq 2$, satisfy the RP condition (7).

(iii) The certainty equivalent $CE_c(\theta, a, \phi)$ of contract c does not depend on δ. Furthermore, $CE_c(\theta, a, \phi)$ is ϵ-supermodular in (c, θ) and (c, a) in the sense that for every $\epsilon > 0$ there exists $\eta = \eta(\epsilon) > 0$ such that

$$CE_{c'}(\theta', a, \phi) - CE_c(\theta', a, \phi) - CE_{c'}(\theta, a, \phi) + CE_c(\theta, a, \phi) > 0 \quad \text{(S.18)}$$

$$CE_{c'}(\theta, a', \phi) - CE_c(\theta, a', \phi) - CE_{c'}(\theta, a, \phi) + CE_c(\theta, a, \phi) > 0 \quad \text{(S.19)}$$

for all $c' > c$, $\theta' \geq \theta + \epsilon$, $a' \geq a + \epsilon$, $(\theta, a) \in [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}]$ and $|\phi| < \eta$ whenever $CE_c(\cdot, \cdot, \cdot)$ is continuous in (θ, a, ϕ).[13]

Proof of Lemma C.4: (i) As is well known, the CARA utility with parameter $a > 0$ is a special case of the HARA utility by setting $\beta = [1 - (1/\alpha)]^{-1/\alpha}$, $\gamma = a[1 - (1/\alpha)]^{1-(1/\alpha)}$ and letting $\alpha \to \pm\infty$. Indeed, upon substituting, we obtain after some algebra

$$U(x; \alpha, \beta, \gamma) = -\left[1 - \frac{ax}{\alpha}\right]^{\alpha} \longrightarrow -\exp(-ax)$$

as $\alpha \to \pm\infty$. To avoid dealing with a divergence to infinity and to obtain the CARA utility, we reparameterize the HARA utility function when $1/\alpha < 1$ and $\beta > 0$ by letting $\phi = 1/\alpha$,

[13]Supermodularity in (c, θ) and (c, a) replaces $\theta' \geq \theta + \epsilon$ and $a' \geq a + \epsilon$ for any $\epsilon > 0$ by $\theta' > \theta$ and $a' > a$, respectively. See footnote 5.

$a = \gamma/\{\beta[1-(1/\alpha)]\}$ and $\delta = [1-(1/\alpha)]\beta^\alpha$. Hence

$$U(x;\alpha,\beta,\gamma) = U(x;a,\phi,\delta) = -\delta(1-\phi a x)^{1/\phi}$$

for $a > 0$, $\phi < 1$, $\delta > 0$ and $1-\phi a x > 0$. Because it is a HARA utility, $U(\cdot;a,\phi,\delta)$ is increasing and concave on $(-\infty, 1/(\phi a))$ if $\phi > 0$ and on $(1/(\phi a), +\infty)$ if $\phi < 0$. In particular, the CARA utility with parameter a is obtained when $(\phi,\delta) \to (0,1)$.

(ii) Fix $\phi < 1$ and $\delta > 0$. Let $a' > a$. We show that $U(x;a',\phi,\delta) = q[U(x;a,\phi,\delta)]$ for some increasing and concave function $q(\cdot)$. See footnote 5. We have

$$U(x;a',\phi,\delta) = -\delta\left(1-\frac{a'}{a}\phi a x\right)^{1/\phi} = -\delta\left\{1-\frac{a'}{a}\left[1-\left(\frac{-U(x;a,\phi,\delta)}{\delta}\right)^{\phi}\right]\right\}^{1/\phi}$$

using the definition of $U(x;a,\phi,\delta)$, which is negative. Thus,

$$q(u) = -\delta\left\{1-\frac{a'}{a}\left[1-\left(\frac{-u}{\delta}\right)^{\phi}\right]\right\}^{1/\phi}.$$

Hence, after some algebra its first derivative is

$$q'(u) = \frac{a'}{a}\left(\frac{-u}{\delta}\right)^{\phi-1}\left\{1-\frac{a'}{a}\left[1-\left(\frac{-u}{\delta}\right)^{\phi}\right]\right\}^{(1/\phi)-1} > 0.$$

Thus, $q(.)$ is increasing as desired. Its second derivative is

$$\begin{aligned} q''(u) &= \frac{a'}{a}\left(\frac{-u}{\delta}\right)^{\phi-2}\left\{1-\frac{a'}{a}\left[1-\left(\frac{-u}{\delta}\right)^{\phi}\right]\right\}^{(1/\phi)-2} \\ &\quad\times\left(-\frac{\phi-1}{\delta}\left\{1-\frac{a'}{a}\left[1-\left(\frac{-u}{\delta}\right)^{\phi}\right]\right\}+\left(\frac{1}{\phi}-1\right)\frac{a'}{a}\left(\frac{-\phi}{\delta}\right)\left(\frac{-u}{\delta}\right)^{\phi}\right) \\ &= \frac{1-\phi}{\delta}\left(1-\frac{a'}{a}\right)\frac{a'}{a}\left(\frac{-u}{\delta}\right)^{\phi-2}\left\{1-\frac{a'}{a}\left[1-\left(\frac{-u}{\delta}\right)^{\phi}\right]\right\}^{(1/\phi)-2} < 0 \end{aligned}$$

since $0 < a < a'$, $\phi < 1$ and $\delta > 0$. Thus, $q(\cdot)$ is concave as desired.

(iii) It follows from (S.6) that the certainty equivalent $CE_c(\theta,a,\phi)$ of contract c does not depend on δ when the utility is $U(x;a,\phi,\delta) = -\delta(1-\phi a x)^{1/\phi}$. Let $\Delta_{c'c}(\theta,a,\phi) \equiv CE_{c'}(\theta,a,\phi) - CE_c(\theta,a,\phi)$. Thus (S.18) and (S.19) become

$$\begin{aligned} \Delta^2_{c'c}(\theta',\theta,a,\phi) &\equiv \Delta_{c'c}(\theta',a,\phi) - \Delta_{c'c}(\theta,a,\phi) > 0 \\ \Delta^2_{c'c}(\theta,a',a,\phi) &\equiv \Delta_{c'c}(\theta,a',\phi) - \Delta_{c'c}(\theta,a,\phi) > 0. \end{aligned}$$

In particular, $CE_c(\theta,a,0)$ is the certainty equivalent of contract c when $U(\cdot;a,\phi,\delta)$ is the CARA utility. It follows from Lemma C.3 that (S.18) and (S.19) hold for all $c' > c$, $\theta' > \theta$, $a' > a$ and $(\theta,a) \in [\underline{\theta},\overline{\theta}] \times [\underline{a},\overline{a}]$ when $\phi = 0$.

Consider (S.18). For $\epsilon > 0$ small, let $\mathcal{K}_\epsilon \equiv \{(\theta', \theta) \in [\underline{\theta}, \overline{\theta}]^2 : \theta' \geq \theta + \epsilon\}$. Also, consider $[\underline{\phi}, \overline{\phi}]$ with $-\infty < \underline{\phi} < 0 < \overline{\phi} < 1$. Fix $c' > c$. Because $CE_c(\theta, a, \phi)$ is continuous in $(\theta, a, \phi) \in [\underline{\theta}, \overline{\theta}] \times [\underline{a}, \overline{a}] \times [\underline{\phi}, \overline{\phi}]$, then $\Delta^2_{c'c}(\theta', \theta, a, \phi)$ is continuous and hence uniformly continuous in $(\theta', \theta, a, \phi) \in \mathcal{K}_\epsilon \times [\underline{a}, \overline{a}] \times [\underline{\phi}, \overline{\phi}]$. Thus, for every $\tilde{\epsilon} > 0$ there exists $\eta = \eta(\epsilon, \tilde{\epsilon})$ independent of (θ', θ, a) such that

$$|\phi| < \eta \quad \Rightarrow \quad |\Delta^2_{c'c}(\theta', \theta, a, \phi) - \Delta^2_{c'c}(\theta', \theta, a, 0)| < \tilde{\epsilon}$$

for all $(\theta', \theta, a) \in \mathcal{K}_\epsilon \times [\underline{a}, \overline{a}]$. But $\Delta^2_{c'c}(\theta', \theta, a, 0)$ is positive and continuous in $(\theta', \theta, a) \in \mathcal{K}_\epsilon \times [\underline{a}, \overline{a}]$. Thus, $\Delta^2_{c'c}(\cdot, \cdot, \cdot, 0) > \underline{m}$ on $\mathcal{K}_\epsilon \times [\underline{a}, \overline{a}]$ for some $\underline{m} = \underline{m}(\epsilon) > 0$. Hence, letting $\tilde{\epsilon} = \underline{m}/2$ there exists $\eta = \eta(\epsilon)$ independent of (θ', θ, a) such that

$$\Delta^2_{c'c}(\theta', \theta, a, \phi) = \Delta^2_{c'c}(\theta', \theta, a, 0) + \left[\Delta^2_{c'c}(\theta', \theta, a, \phi) - \Delta^2_{c'c}(\theta', \theta, a, 0)\right] > \underline{m}/2 > 0$$

for all $(\theta', \theta, a) \in \mathcal{K}_\epsilon \times [\underline{a}, \overline{a}]$ and $|\phi| < \eta$. Because there is a finite number of pairs (c, c') such that $c' > c$, this argument shows that (S.18) also holds for all $c' > c$. A similar argument with $\mathcal{K}_\epsilon \equiv \{(a', a) \in [\underline{a}, \overline{a}]^2 : a' \geq a + \epsilon\}$ establishes (S.19). □

Lemma C.5: *Let the coverages (t_c, dd_c), $c = 0, 1, \ldots, C \geq 2$ satisfy the RP condition (7). The certainty equivalent $CE_c(\theta, a)$ associated with coverage c satisfies condition (S.10) when $U(\cdot; a)$ and $P(\cdot; \theta)$ are the CARA utility and the Poisson distribution, respectively.*

Proof of Lemma C.5: From the proof of Lemma C.3, we have

$$\begin{aligned}
\frac{\partial CE_c(\theta, a)}{\partial \theta} - \frac{\partial CE_{c+1}(\theta, a)}{\partial \theta} &= -\frac{1}{a}[\phi_a(dd_c) - \phi_a(dd_{c+1})] = -\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD, \\
\frac{\partial CE_c(\theta, a)}{\partial a} - \frac{\partial CE_{c+1}(\theta, a)}{\partial a} &= -\theta \frac{\partial}{\partial a}\left[\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD\right].
\end{aligned}$$

Thus upon simplifying and rearraging terms, condition (S.10) is equivalent to

$$\frac{\partial}{\partial a}\left[\log \int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD\right] > \frac{\partial}{\partial a}\left[\log \int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD\right],$$

which is equivalent to

$$\frac{\partial}{\partial a}\left[\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD \Big/ \int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD\right] > 0,$$

i.e., that the ratio within brackets is increasing in a. Because

$$\frac{\int_{dd_{c+2}}^{dd_c} e^{aD}[1 - H(D)]dD}{\int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD} = 1 + \frac{\int_{dd_{c+1}}^{dd_c} e^{aD}[1 - H(D)]dD}{\int_{dd_{c+2}}^{dd_{c+1}} e^{aD}[1 - H(D)]dD},$$

it follows that condition (S.10) is equivalent to the LHS increasing in a. This is true because $\int_{dd_{c+2}}^{dd} e^{aD}[1-H(D)]dD$ is log-supermodular in (a, dd) by Lemma C.1 upon letting $x = a$, $y = dd$ and $y_\dagger = dd_{c+2}$.

Lemma C.6: *Let J and J' be random variables distributed as $P(\cdot)$ and $P'(\cdot)$ on $I\!N \equiv \{0, 1, 2, \ldots\}$. Let $X_0 \equiv 0$ and $X_1, X_2, \ldots$ be i.i.d. random variables independent of (J, J') with support $(0, \overline{x})$, $0 < \overline{x} \le +\infty$. Suppose that $P'(\cdot) \overset{FOSD}{\succ} P(\cdot)$. Then $\sum_{i=0}^{J'} X_i \overset{FOSD}{\succ} \sum_{i=0}^{J} X_i$.*

Proof of Lemma C.6: By definition, $\sum_{i=0}^{J'} X_i \overset{FOSD}{\succ} \sum_{i=0}^{J} X_i$ is equivalent to

$$\Pr\left[\sum_{i=0}^{J} X_i \le x\right] \ge \Pr\left[\sum_{i=0}^{J'} X_i \le x\right] \tag{S.20}$$

for all $x \ge 0$ with strict inequality for some $x > 0$. Let $F_{S_j}(\cdot)$ be the cdf of $S_j \equiv \sum_{i=0}^{j} X_i = \sum_{i=1}^{j} X_i$ for $j \in I\!N$. Note that $F_{S_0}(x) = 1$ for $x \ge 0$. Thus, we have

$$\Pr\left[\sum_{i=0}^{J} X_i \le x\right] = \mathrm{E}\left\{\Pr\left[\sum_{i=0}^{J} X_i \le x | J\right]\right\} = P(0) + \sum_{j=1}^{+\infty} F_{S_j}(x)P(j) = \sum_{j=0}^{+\infty} F_{S_j}(x)P(j)$$

for $x \ge 0$. A similar expression holds when J is replaced by J' with $P(j)$ replaced by $P'(j)$. Thus (S.20) is equivalent to

$$\sum_{j=0}^{+\infty} F_{S_j}(x)P(j) \ge \sum_{j=0}^{+\infty} F_{S_j}(x)P'(j).$$

Let $U(j; x) \equiv -F_{S_j}(x)$ for $j \in I\!N$ and $x \ge 0$. Hence we want to show that

$$\mathrm{E}[U(J'; x)] \ge \mathrm{E}[U(J; x)] \tag{S.21}$$

for all $x \ge 0$ with strict inequality for some $x > 0$ whenever $P'(\cdot) \overset{FOSD}{\succ} P(\cdot)$.

To this end, we show that $U(\cdot; x)$ is nondecreasing on $I\!N$ and increasing on $(0, \overline{x})$. To see this, we note that $F_{S_{j+1}}(\cdot) \overset{FOSD}{\succ} F_{S_j}(\cdot)$ for $j \ge 0$ because $S_{j+1} = S_j + X_{j+1}$ with $X_{j+1} \ge 0$ and $\overline{x} > 0$. Specifically, for $j \in I\!N$ and $x \ge 0$, we have $F_{S_j}(x) \ge F_{S_{j+1}}(x)$ with strict inequality for $\{j = 0$ and $x \in [0, \overline{x})\}$ or $\{j \ge 1$ and $x \in (0, (j+1)\overline{x})\}$. Hence $U(j; x)$ is nondecreasing in $j \in I\!N$ for every $x \ge 0$. Moreover, it is easy to verify that $U(\cdot; x)$ is increasing on $I\!N$ when $x \in (0, \overline{x})$. Now, because $P'(\cdot) \overset{FOSD}{\succ} P(\cdot)$, it follows that (S.21) holds for all $x \ge 0$ with strict inequality for $x \in (0, \overline{x})$. $\square$

References

Aryal, Gaurab, **Isabelle Perrigne**, and **Quang Vuong** (2012): "Nonidentification of Insurance Models with Probability of Accidents," Unpublished manuscript, University of Virginia.

Athey, Susan (2002): "Monotone Comparative Statics under Uncertainty," *Quarterly Journal of Economics*, 117, 187-223.

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

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

Gollier, Christian (2001): *The Economics of Risk and Time*, MIT Press: Cambridge.

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

Johnson, Norman L., **Adrienne W. Kemp** and **Samuel Kotz** (2005): *Univariate Discrete Distributions*, 3rd Edition, New Jersey, USA: John Wiley and Sons.

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

Milgrom, Paul and **John Roberts** (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255-1277.

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

Rothschild, Michael and **Joseph Stiglitz** (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225-243.

Teicher, Henry (1961): "Identifiability of Mixtures," *Annals of Mathematical Statistics*, 32, 244-248.

Topkis, Donald (1978): "Minimizing a Submodular Function on a Lattice," *Operation Research*, 26, 305-321.

Vives, Xavier (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305-321.