Discrete Choice under Risk with Limited Consideration*

Levon Barseghyan[†] Francesca Molinari[‡] Matthew Thirkettle[§]

February 18, 2021

For Online Publication

*We are grateful to Liran Einav, three anonymous referees, Abi Adams, Jose Apesteguia, Miguel Ballester, Arthur Lewbel, Chuck Manski, and Jack Porter for useful comments and constructive criticism. For comments and suggestions we thank the participants to the 2017 Barcelona GSE Summer Forum on Stochastic Choice, the 2018 Cornell Conference "Identification and Inference in Limited Attention Models", the 2018 Penn State-Cornell Conference on Econometrics and IO, the 2018 GNYMA Conference, the 2019 ASSA meetings, the IFS 2019 "Consumer Behaviour: New Models, New Methods" Conference, and to seminars at Stanford, Berkeley, UCL, Wisconsin, Bocconi, and Duke. Part of this research was carried out while Barseghyan and Molinari were on sabbatical leave at the Department of Economics at Duke University, whose hospitality is gratefully acknowledged. We gratefully acknowledge support from National Science Foundation grant SES-1824448 and from the Institute for Social Sciences at Cornell University.

[†]Department of Economics, Cornell University, lb247@cornell.edu.

[‡]Department of Economics, Cornell University, fm72@cornell.edu.

[§]Department of Economics, Rice University, mt77@rice.edu.

Appendices

A Proofs

Proof of Fact 4. If $c_{1,2}(x)$ is less than $c_{2,3}(x)$, then, for any DM with preference ν s.t. $c_{1,2}(x) < \nu < c_{2,3}(x)$, d_2 is preferred to both d_1 and d_3, i.e. we are in Case (i). If $c_{1,2}(x) > c_{2,3}(x)$, then d_2 is either dominated by either d_1 or d_3. The relative location of $c_{1,3}(x)$ is established as follows. First, suppose $c_{1,3}(x) < c_{1,2}(x) < c_{2,3}(x)$. For any $\nu \in (c_{1,3}(x), c_{1,2}(x))$ we have $U_\nu(d_3, x) > U_\nu(d_1, x) > U_\nu(d_2, x) > U_\nu(d_3, x)$, which is an obvious contradiction. Second, suppose $c_{2,3}(x) < c_{1,2}(x) < c_{1,3}(x)$. Then, for any $\nu \in (c_{1,2}(x), c_{1,3}(x))$ we have $U_\nu(d_1, x) > U_\nu(d_3, x) > U_\nu(d_2, x) > U_\nu(d_1, x)$, which is an obvious contradiction. The remaining two possibilities are excluded following the same logic. □

We maintain that x has strictly positive density on $\mathcal{S}$ (Assumption T0), its density is continuous (Assumption T1), and that preferences are continuous and strictly monotone in x. Therefore, if x is a scalar and $c_{1,2}(x)$ covers $[\nu^l, \nu^u]$, it is sufficient to consider an interval $[x^l, x^u] \subset \mathcal{S}$ such that $[\nu^l, \nu^u] = \{c_{1,2}(x) : x \in [x^l, x^u]\}$.

The following lemma is useful for establishing Theorem 1.

Lemma A.1. *Suppose Assumptions T0-T2 and I1 hold. Suppose* $c_{1,2}(x) < c_{1,j}(x)$, $\forall x \in \mathcal{X}$. *Let* $\{x^t\}_{t=1}^\infty$ *be s.t.* $c_{1,2}(x^t) = c_{1,j}(x^{t+1})$, $x^t \in \mathcal{X}$. *Then* $\exists T < \infty$ *s.t.* $c_{1,2}(x^T) < 0$.

Proof. The cutoff $\{c_{1,2}(x^t)\}_{t=1}^\infty$ is a strictly declining sequence. Suppose all its elements are non-negative. Then it converges to some $\nu^\infty \geq 0$ such that $\nu^\infty = c_{1,2}(x^\infty) = c_{1,j}(x^\infty)$ for some $x^\infty \in \mathcal{X}$, a contradiction. □

Proof of Theorem 1. The second condition in the theorem implies that the cutoffs are ordered: $c_{1,j}(x) < c_{1,j+1}(x)$ for all $x \in \mathcal{X}$. Hence

$$\begin{aligned}
\Pr(d = d_1|x) &= \sum_{j=2}^{D} \sum_{\substack{\mathcal{K}\subset\mathcal{D}: \\ 1,j\in\mathcal{K}, \\ 2,\ldots,j-1\notin\mathcal{K}}} Q(\mathcal{K})F(c_{1,j}(x)) + Q(\{d_1\}) \\
&= \sum_{j=2}^{D} \mathcal{O}(\{d_1, d_j\}; \{d_2, \ldots, d_{j-1}\})F(c_{1,j}(x)) + \mathcal{O}(d_1; \emptyset) \\
&\equiv \sum_{j=2}^{D} \Lambda_j F(c_{1,j}(x)) + \Lambda_1,
\end{aligned}$$

so that

$$\frac{d\Pr(d = d_1|x)}{dx} = \sum_{j=2}^{D} \Lambda_j f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx}.$$

By Assumption I1, we can set $x^u = x^{\bar{\nu}}$ s.t. $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$ and similarly $x^l = x^0$ s.t. $c_{1,2}(x^0) = 0$. It may be the case that $c_{1,\hat{D}}(x^0) < \bar{\nu}$ and $c_{1,\hat{D}+1}(x^0) > \bar{\nu}$ for some $\hat{D} \geq 2$. Then, $\forall j > \hat{D}$, Λ_j does not enter the expression for the derivative of $\Pr(d = d_1|x)$, $\forall x \in [x^0, x^{\bar{\nu}}]$, because $f(c_{1,j}(x)) = 0$. Henceforth, we only consider the relevant alternatives for the derivative of $\Pr(d = d_1|x)$, namely $j \leq \hat{D}$.

Next, consider the derivative of $\Pr(d = d_j|x)$. By Fact 3, the term Λ_j is the leading coefficient on $f(\cdot)$ for this derivative. There exists $x^j \in \mathcal{X}$ such that $c_{1,j}(x^j) = \bar{\nu}$. Thus,

$$\lim_{x \nearrow x^j} \frac{d\Pr(d = d_j|x)}{dx} = -\Lambda_j f(\bar{\nu}) \frac{dc_{1,j}(x^j)}{dx}, \quad \forall j : 2 \leq j \leq \hat{D}.$$

The ratio of $\lim\limits_{x \nearrow x^j} d\Pr(d = d_j|x)/dx$ and $\lim\limits_{x \nearrow x^2} d\Pr(d = d_2|x)/dx$ identifies $\Omega_j \equiv \frac{\Lambda_j}{\Lambda_2}$, where $\Lambda_2 \neq 0$ by the first assumption in the theorem. Rewrite the derivative of $\Pr(d = d_1|x)$ as follows:

$$\begin{aligned}
\frac{d\Pr(d=d_1|x)}{dx} &= \sum_{j=2}^{\hat{D}} \Lambda_j f(c_{1,j}(x))\frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \frac{\Lambda_j}{\Lambda_2}\Lambda_2 f(c_{1,j}(x))\frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \Omega_j[\Lambda_2 f(c_{1,j}(x))]\frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \Omega_j \hat{f}(c_{1,j}(x))\frac{dc_{1,j}(x)}{dx},
\end{aligned}$$

where $\hat{f}(\nu) \equiv \Lambda_2 f(c_{1,j}(x))$. Equipped with Ω_j, we can recover $\hat{f}(\nu)$ sequentially. Note that $\forall x$ s.t. $c_{1,2}(x) \leq \bar{\nu}$ and $c_{1,3}(x) > \bar{\nu}$, the up-to-scale density $\hat{f}(c_{1,2}(x))$ is identified. Indeed, it is the only unknown in the expression above. We proceed as follows.

First, let x^1 be such that $c_{1,3}(x^1) = \bar{\nu}$. Then, $\hat{f}(\cdot)$ is identified on $[c_{1,2}(x^1), \bar{\nu}]$.

Second, let x^2 be such that $c_{1,2}(x^2) = c_{1,3}(x^1)$. Now $\hat{f}(\nu)$ is identified on $[c_{1,2}(x^2), \bar{\nu}]$ because in the expression for the derivative of $\Pr(d = d_1|x)$ all cutoffs $c_{1,j}(x)$, $j > 2$, lie on the part of the support where the up-to-scale density is known.

Identification of $\hat{f}(\nu)$ on $[0, \bar{\nu}]$ attains by repeating the above step. Indeed, by Lemma A.1 $c_{1,2}(x^t)$ reaches the lower end of the support in a finite number of steps. Finally, the scale is recovered by integrating $\hat{f}(\nu)$ over its support:

$$\Lambda_2 = \Lambda_2 \int_0^{\bar{\nu}} f(\nu)d\nu = \int_0^{\bar{\nu}} \hat{f}(\nu)d\nu.$$

Therefore $f(\cdot)$ is identified, as required. The term $\mathcal{O}(d_1;\emptyset) = \Lambda_1 = \Pr(d = d_1|x^{\bar{\nu}})$ is also identified, and so are $\mathcal{O}(d_1, d_2;\emptyset) = \Lambda_2$ and $\mathcal{O}(\{d_1, d_j\};\{d_2, \ldots, d_{j-1}\}) = \Lambda_j$.

□

Proof of Theorem 2. The second condition of Theorem 1 implies that the cutoffs are ordered: $c_{1,j}(x) < c_{1,j+1}(x)$ for all $x \in \mathcal{X}$. Hence,

$$\frac{d\Pr(d=d_1|x)}{dx} = \sum_{j=2}^{D} \Lambda_j(c_{1,j}(x)) f(c_{1,j}(x))\frac{dc_{1,j}(x)}{dx},$$

where

$$\Lambda_j(\nu) = \begin{cases} \underline{\Lambda}_j \equiv \sum_{\substack{\mathcal{K}\subset\mathcal{D}:\\ 1,j\in\mathcal{K},\\ 2,\ldots,j-1\notin\mathcal{K}}} \underline{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu < \nu^* \\ \overline{\Lambda}_j \equiv \sum_{\substack{\mathcal{K}\subset\mathcal{D}:\\ 1,j\in\mathcal{K},\\ 2,\ldots,j-1\notin\mathcal{K}}} \overline{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu \geq \nu^* \end{cases}.$$

Similar to the proof of Theorem 1, we only consider the relevant alternatives for the derivative of $\Pr(d = d_1|x)$, namely $j \leq \hat{D}$.

We start at $x^{\bar{\nu}}$ and hence $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$. As we lower x we check whether $\frac{d\Pr(d=d_1|x)}{dx}$ jumps. If it does not, identification of $f(\cdot)$ attains by the proof of Theorem 1.

Suppose there is a point of discontinuity. It arises when a cutoff $c_{1,j}(x)$ crosses the breakpoint ν^*. The identity of the cutoff and hence $\nu^* = c_{1,j}(x)$ is identified by the fact there is a unique $\frac{d\Pr(d=d_j|x)}{dx}$ that also jumps. Equipped with the identity of ν^* the proof proceeds similarly to that of Theorem 1. Indeed, all $\overline{\Omega}_j \equiv \overline{\Lambda}_j/\overline{\Lambda}_2$ are identified and so is $\hat{f}(\nu) \equiv \overline{\Lambda}_2 f(\nu)$ for all $\nu > \nu^*$.

The additional step is how to identify $\underline{\Lambda}_j$ and $\underline{\Omega}_j \equiv \underline{\Lambda}_j/\underline{\Lambda}_2$. Start with $\underline{\Lambda}_2$. Consider an x^* s.t. $c_{1,2}(x^*) = \nu^*$. The derivatives $\frac{d\Pr(d=d_1|x)}{dx}$ from the left and from the right of x^* identify $\underline{\Lambda}_2 f(\nu^*)$ and $\overline{\Lambda}_2 f(\nu^*)$. Hence, the ratio $\underline{\Lambda}_2/\overline{\Lambda}_2$ is identified. Exactly the same logic applies to all other $\underline{\Lambda}_j$'s whenever $c_{1,j}(x)$ crosses ν^*. We can then rewrite

$$\frac{d\Pr(d = d_1|x)}{dx} = \sum_{j=2}^{\hat{D}} \left(\overline{\Omega}_j\right)^{\mathbb{1}(c_{1,j}(x)\geq\nu^*)} \left(\underline{\Omega}_j \frac{\underline{\Lambda}_2}{\overline{\Lambda}_2}\right)^{\mathbb{1}(c_{1,j}(x)<\nu^*)} \hat{f}(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},$$

Now all coefficients of $\hat{f}(c_{1,j}(x))$ are identified, and identification of $\hat{f}(\cdot)$ proceeds to the left of ν^*. Once it is identified, we integrate it over the support to recover $\overline{\Lambda}_2$. Hence $\underline{\Lambda}_2$ and $f(\cdot)$ are identified.

When the breakpoint occurs at x^* rather than ν^*, the same proof strategy can be applied.

□

Proof of Proposition 1. The second condition in the theorem implies that $c_{1,2}(x) < c_{1,j}(x)$ for any $j > 2$. The cutoffs $c_{1,j}(x)$'s, $j > 2$, are irrelevant for evaluating $\Pr(d = d_1|x)$ by the first condition of the theorem. Therefore,

$$\frac{d\Pr(d=d_1|x)}{dx} = \left(\sum_{\mathcal{K}\subset\mathcal{D}:1,2\in\mathcal{K}} Q(\mathcal{K})\right) f(c_{1,2}(x))\frac{dc_{1,2}(x)}{dx} \equiv \alpha f(c_{1,2}(x))\frac{dc_{1,2}(x)}{dx}.$$

Consequently, the product $\alpha f(c_{1,2}(x))$ can be written in terms of data:

$$\alpha f(c_{1,2}(x)) = \frac{\frac{d\Pr(d=d_1|x)}{dx}}{\frac{dc_{1,2}(x)}{dx}},$$

and hence variation in x guarantees that $f(\nu)$ is identified up-to-scale on $[c_{1,2}(x^l), c_{1,2}(x^u)] = [\nu^l, \nu^u]$. It follows that $F(\nu|\nu \in [\nu^l, \nu^u])$ is identified. □

Proof of Theorem 3. Under Condition *1* of the Theorem there can only be three types of consideration sets. The first type are all possible subsets of $\{d_1, \ldots, d_j\}$; the second type are all possible subsets of $\{d_{j+1}, \ldots, d_D\}$. The third type necessarily contains both d_j and d_{j+1}. The probability of choosing an alternative in $\{d_1, \ldots, d_j\}$ is one for the first type and zero for the second type. Hence,

$$\frac{d\Pr(d\in\{d_1,\ldots,d_j\}|x)}{dx} = \sum_{\mathcal{K}\subset\mathcal{D}:j,j+1\in\mathcal{K}} Q(\mathcal{K}) f(c_{j,j+1}(x))\frac{dc_{j,j+1}(x)}{dx}.$$

The rest of the proof follows the same steps as in the proof of Proposition 1, except we now track $c_{j,j+1}(x)$. □

Proof of Theorem 4. Let ν, $\tilde{x}$, $\mathcal{N}_\epsilon(\tilde{x}) \equiv \{x : \|x - \tilde{x}\| < \epsilon\}$ satisfy Condition *2* in the theorem. Then, $\nu = c_{j,k}(\tilde{x})$ for all j, k. Consider any pair of alternatives (d_j, d_k). Since utility is strictly monotone in x_j and continuous, for each $\mathcal{L} \subseteq \mathcal{D} \setminus \{j, k\}$ we can find $x \in \mathcal{N}_\epsilon(\tilde{x})$ such that

$$U_\nu(d_j, x) = U_\nu(d_k, x); \tag{A.1}$$

$$U_\nu(d_l, x) > U_\nu(d_j, x) \quad \forall l \in \mathcal{L}; \tag{A.2}$$

$$U_\nu(d_j, x) > U_\nu(d_l, x) \quad \forall l \in \mathcal{D} \setminus \{\mathcal{L} \cup \{j, k\}\}; \tag{A.3}$$

The remainder of the proof proceeds in two steps.

Step 1: Identification of $f(\nu)\mathcal{Q}_\nu(\mathcal{K})$: The singleton sets occur with zero probability by Condition *1* in the theorem, so it remains to show identification for consideration sets larger than one. Consider any two alternatives (d_j, d_k). We claim that the following statement holds for $n = 0, \ldots, D$:

$P(n)$: For all $\mathcal{K} \subset \mathcal{D} \setminus \{j,k\}$ satisfying $|\mathcal{K}| \leq n$, the quantity $f(\nu)\mathcal{Q}_\nu(\{j,k\} \cup \mathcal{K})$ is identified.

To show this for $P(0)$, set $\mathcal{L} = \mathcal{D} \setminus \{j,k\}$. In this case $\mathcal{K} = \emptyset$. Let x satisfy Equations (A.1)-(A.3). Then, all alternatives d_l, $l \neq j,k$, are preferred to d_j and d_k at ν and $c_{j,k}(x) = \nu$. Hence:

$$\frac{\frac{\partial \Pr(d=d_j|x)}{\partial x_k}}{\frac{\partial c_{j,k}(x)}{\partial x_k}} = f(\nu)\mathcal{Q}_\nu(\{j,k\}).$$

It follows that $f(\nu)\mathcal{Q}_\nu(\{j,k\})$ is identified.

Next, suppose $P(n-1)$ is true. Consider any $\mathcal{K} \subset \mathcal{D} \setminus \{j,k\}$ such that $|\mathcal{K}| = n$. Let $\mathcal{L} = \mathcal{D} \setminus (\mathcal{K} \cup \{j,k\})$. Let x satisfy Equations (A.1)-(A.3). Then,

$$\begin{aligned}\frac{\frac{\partial \Pr(d=d_j|x)}{\partial x_k}}{\frac{\partial c_{j,k}(x)}{\partial x_k}} &= f(\nu) \sum_{\mathcal{C} \subset \mathcal{K}} \mathcal{Q}_\nu(\{j,k\} \cup \mathcal{C}) \\ &= f(\nu)\mathcal{Q}_\nu(\{j,k\} \cup \mathcal{K}) + \sum_{\mathcal{C} \subset \mathcal{K} : |\mathcal{C}| < n} f(\nu)\mathcal{Q}_\nu(\{j,k\} \cup \mathcal{C}).\end{aligned}$$

The LHS of this expression is known, and the second term on the RHS is identified by the induction step. Therefore $P(n)$ holds.

Since d_j and d_k were chosen arbitrarily, it follows that $f(\nu)\mathcal{Q}_\nu(\mathcal{K})$ is identified for all $\mathcal{K} \subset \mathcal{D}$.

Step 2: Identification of $f(\nu)$ and $\mathcal{Q}_\nu(\mathcal{K})$. Since

$$\sum_{\mathcal{K} \subset \mathcal{D}} f(\nu)\mathcal{Q}_\nu(\mathcal{K}) = f(\nu) \sum_{\mathcal{K} \subset \mathcal{D}} \mathcal{Q}_\nu(\mathcal{K}) = f(\nu),$$

$f(\nu)$ is identified. Identification of $\mathcal{Q}_\nu(\mathcal{K})$ follows from Step 1. □

Proof of Corollary 1. The proof follows the same steps as the proof of Theorem 4, but with the following two modifications:

First modification: In Step 1 in the proof of Theorem 4, we start with $d_j = d_1$ and loop over $d_k \in \{d_2, \ldots, d_D\}$. This ensures that we only take derivatives with respect to x_k, $k > 1$. Hence, $f(\nu)\mathcal{Q}_\nu^{x_1}(\mathcal{K})$ is identified for all sets $\mathcal{K} \subset \mathcal{D} : |\mathcal{K}| > 1$.

Second modification: In Step 2 we obtain

$$f(\nu)\mathcal{Q}^{x_1}(d_1) + \sum_{\mathcal{K}\subset\mathcal{D}:|\mathcal{K}|>1} f(\nu)\mathcal{Q}^{x_1}_{\nu}(\mathcal{K}) = f(\nu)\sum_{\mathcal{K}\subset\mathcal{D}} \mathcal{Q}^{x_1}_{\nu}(\mathcal{K}) = f(\nu).$$

Since the second term on the LHS is known, $f(\nu)(1-\mathcal{Q}^{x_1}(d_1))$ is identified for all $\nu \in [0,\bar{\nu}]$. The scale is identified, because

$$(1-\mathcal{Q}^{x_1}(d_1)) = \int_0^{\bar{\nu}} f(\nu)(1-\mathcal{Q}^{x_1}(d_1))d\nu.$$

Once the scale is identified, $f(\nu)$ is identified and so are $\mathcal{Q}^{x_1}_{\nu}(\mathcal{K})$, $\forall\mathcal{K}\subset\mathcal{D}$. $\square$

Proof of Proposition 2. For the purpose of obtaining a contradiction, suppose that there is full consideration. Then

$$\begin{aligned}
\Pr(d\in\mathcal{L}|x) &= \Pr\left(\arg\max_{j\in\mathcal{D}} U_{\nu}(d_j,x)\in\mathcal{L}\middle| x\right)\\
&= \Pr(\nu\in[0,\nu^*))\\
&= F(\nu^*)\\
&= \Pr\left(\arg\max_{j\in\mathcal{D}} U_{\nu}(d_j,x')\in\mathcal{L}'\middle| x'\right)\\
&= \Pr(d\in\mathcal{L}'|x').
\end{aligned}$$

This is a contradiction. Therefore there is limited consideration.

$\square$

The following two Lemmas are used in the proof of Theorem 5. The proofs of these Lemmas rest on the following claims.

1. The probability of alternative d_j being chosen can only increase in its consideration probability.

2. The probability of alternative d_j being chosen can only decline in consideration probability of any other alternative d_k.

3. The probability of alternative d_j or d_k being chosen can only increase in the consideration probability of d_j as the positive effect of this change on $Pr(d=d_j|x)$ dominates the negative effect on $Pr(d=d_k|x)$.

4. The probability of an alternative in $\mathcal{K}$ being chosen can only decline in consideration probability of any alternative that does not belong to $\mathcal{K}$.

Lemma A.2. *Consider the Basic ARC model. For any $\mathcal{K} \subset \mathcal{D}$, $\sum_{j \in \mathcal{K}} \Pr(d = d_j|x)$ is increasing in φ_k, $\forall k \in \mathcal{K}$, and decreasing in φ_k, $\forall k \notin \mathcal{K}$.*

Proof. Fix $\mathcal{K}$ and consider any $j \in \mathcal{K}$. For each ν and $l \in \mathcal{K}$, $l \neq j$, either $j \in \mathcal{B}_\nu(d_l, x)$ or not. If $j \notin \mathcal{B}_\nu(d_l, x)$, then $\Pr(d = d_l|x, \nu)$ does not depend on φ_j. Hence,

$$\sum_{l \in \mathcal{K}} \Pr(d = d_l|x, \nu) = A + \Pr(d = d_j|x, \nu) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \Pr(d = d_l|x, \nu),$$

where A is a constant that collects terms that do not depend on φ_j. Continuing,

$$\begin{aligned}
\sum_{l \in \mathcal{K}} \Pr(d = d_l|x, \nu) &= A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x)} (1 - \varphi_k) \\
&= A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l (1 - \varphi_j) \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \\
&= A + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \left(\varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \right) \\
&\quad + \left(\prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) - \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \right) \varphi_j \\
&\equiv \tilde{A} + B\varphi_j.
\end{aligned}$$

Since $\mathcal{B}_\nu(j, x) \subset \mathcal{B}_\nu(d_l, x)$ whenever $j \in \mathcal{B}_\nu(d_l, x)$,

$$B = \left(\prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) \right) \left(1 - \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{\mathcal{B}_\nu(d_j, x), j\}} (1 - \varphi_k) \right) \geq 0.$$

Therefore, $\sum_{j \in \mathcal{K}} \Pr(d = d_j|x) = \int \sum_{j \in \mathcal{K}} \Pr(d = d_j|x, \nu) dF$ is increasing in φ_j.

Finally, for any $k \notin \mathcal{K}$, φ_k may only appear on the RHS as $(1 - \varphi_k)$. Hence $\sum_{j \in \mathcal{K}} \Pr(d = d_j|x)$ is decreasing in φ_k. □

Lemma A.3. *Consider the Basic ARC model. For any $\mathcal{K} \subset \mathcal{D}$, $\sum_{j \in \mathcal{K}} \Pr(d = d_j|x)$ is strictly increasing in φ_j, $j \in \mathcal{K}$, whenever there is an open interval of ν's at which alternative d_j is preferred to all of the always-considered alternatives. It is strictly decreasing in φ_k, $j \notin \mathcal{K}$,* *whenever there is an open interval of ν's and $l \in \mathcal{K}$ such that at these ν's alternative d_k is preferred to d_l and d_l is preferred to all of the always-considered alternatives.*

Proof. To show the first claim, notice that $B = 0$ in the proof of Lemma A.2 if and only if $\varphi_k = 1$ for some $k \in \mathcal{B}_\nu(d_j, x)$.

To show the second claim, consider any $j \in \mathcal{K}$. Then,

$$\Pr(d = d_j|x, \nu) = \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j,x)} (1 - \varphi_k).$$

For this to be strictly decreasing in φ_k, it must be the case that $k \in \mathcal{B}_\nu(d_j, x)$ and $\varphi_l < 1$ for all $l \in \mathcal{B}_\nu(d_j, x) \setminus \{k\}$. □

Proof of Theorem 5. By the proof of Theorem 1, $f(\cdot)$ and $\Lambda_2 = \varphi_1\varphi_2$ are identified. The consideration parameter φ_1 is identified by $\Pr(d = d_1|x^{\bar{\nu}}) = \varphi_1$, where $x^{\bar{\nu}}$ is s.t. $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$. Since Λ_2 is known, φ_2 is also identified. The rest of the proof is about identification of the remaining consideration parameters.

To identify φ_j take an x such that $\Pr(d = d_j|x) \neq 0$. Denote $\mathcal{E} = \{k : \Pr(d = d_k|x) \neq 0\}$. We claim that $\Pr(d = d_k|x), \forall k \in \mathcal{E}$, does not depend on φ_l for $l \notin \mathcal{E}$. Suppose otherwise. That is, suppose there exists d_l such that $\Pr(d = d_l|x) = 0$ and $\Pr(d = d_k|x)$ depends on φ_l for some $k \in \mathcal{E}$. Then, for each ν there is an always-considered alternative that is preferred to d_l. Since $\Pr(d = d_k|x)$ depends on φ_l, there exists $\nu \in [0, \bar{\nu}]$ such that d_l is preferred to d_k. However, the always-considered alternative that is preferred to d_l at ν is also preferred to d_k by transitivity. This leads to a contradiction, because a DM with such preferences will never choose d_k in the first place. Therefore, $\Pr(d = d_k|x)$ does not on φ_l for any $l \notin \mathcal{E}$.

Since $F(\cdot)$ is already identified, $\{\Pr(d = d_k|x)\}_{k\in\mathcal{E}}$ defines a system of $|\mathcal{E}|$ non-linear equations, where the only unknowns are φ_k, $k \in \mathcal{E}$. This system has a unique solution. Suppose to the contrary that two sets of consideration parameters $\{\varphi_k\}_{k\in\mathcal{E}}$ and $\{\varphi'_k\}_{k\in\mathcal{E}}$ solve this system and they are distinct. Denote $\mathcal{E}_+ = \{k : \varphi_k > \varphi'_k\}$. By Lemma A.3, $\sum_{k\in\mathcal{E}_+} \Pr(d = d_k|x)$ is strictly larger at $\{\varphi_k\}_{k\in\mathcal{E}}$ than at $\{\varphi'_k\}_{k\in\mathcal{E}}$. Hence, only one of these sets could satisfy data. Therefore there is a unique set of $\{\varphi_k\}_{k\in\mathcal{E}}$ that solves this system of equations, and φ_j is identified as claimed. □

Proof of Theorem 6. Step 1. Suppose $D = 3$. Then $d^* = 2$ and

$$\Pr(d = d_1|x) = \int_0^{c_{1,2}(x)} \varphi_1(\nu)dF \quad \text{and} \quad \Pr(d = d_3|x) = \int_{c_{2,3}(x)}^{\bar{\nu}} \varphi_3(\nu)dF.$$

The ratio of the derivatives of these two moments yields $\frac{\varphi_1(\nu)}{\varphi_3(\nu)}$, where

$$\begin{aligned}\varphi_1(\nu) &= \varphi_1(1-\alpha(\nu))\\ \varphi_3(\nu) &= \varphi_3(1+\alpha(\nu)),\end{aligned}$$

First, $\frac{\varphi_1}{\varphi_3}$ is identified when x and x' are chosen such that $c_{1,2}(x) = c_{2,3}(x') = \bar{\nu}$. Once $\frac{\varphi_1}{\varphi_3}$ is identified, $\frac{1-\alpha(\nu)}{1+\alpha(\nu)}$ is known for all ν; hence, $\alpha(\nu)$ can be solved for. Identification of $f(\nu)$ follows from substituting $\alpha(\nu)$ into the expression for $\frac{\Pr(d=d_1|x)}{dx}$.

Step 2. Let $D > 3$. We identify d^* at $x^{\bar{\nu}}$. The smallest j such that $\Pr(d = d_j|x^{\bar{\nu}}) = 0$ yields $d^* = j - 1$ (or $d^* = D$ if no such j exists). We return to the case that $d^* = 1$, $d^* = 2$, or $d^* = D$ at the end of this proof.

Step 3. Using large support we establish that φ_D is a decreasing function of φ_1. We have

$$\Pr(d = d_1|x^{\bar{\nu}}) = \varphi_1 \int_0^{\bar{\nu}} (1-\alpha(\nu))dF(\nu) = \varphi_1(1 - E\alpha(\nu)).$$

Similarly,

$$\Pr(d = d_D|x^0) = \varphi_D \int_0^{\bar{\nu}} (1+\alpha(\nu))dF(\nu) = \varphi_D(1 + E\alpha(\nu)).$$

Hence

$$\varphi_1 = \frac{\Pr(d = d_1|x^1)}{2 - \frac{\Pr(d=d_D|x^0)}{\varphi_D}}.$$

Step 4. This is an intermediate step, which we use later in the proof. By Fact 3,

$$c_{1,j}(x) < c_j^*(x) \equiv \min_k \{\{c_{k,j}(x)\}_{1<k<j}, \{c_{j,k}(x)\}_{j<k\leq D}\}, \ \forall j.$$

Moreover any sequence $\{x^s\}_{s=1}^{\infty}$ such that $c_j^*(x^s) = c_{1,j}(x^{s+1})$ will reach the lower bound of the support in finite number of steps. Otherwise, by the argument in the proof of Lemma A.1, $c_j^*(x^s)$ and $c_{1,j}(x^s)$ converge to the same point in the interior of the support, which contradicts the assumptions of the theorem.

Step 5. Identification of $\{\varphi_j\}_{1<j<d^*}$. For each ν in a sufficiently small neighborhood near $\bar{\nu}$,

say $(\bar{\nu}-\varepsilon, \bar{\nu})$, and for each j, there is an x^j such that $c_{1,j}(x^j) = \nu$ and $c_{k,j}(x^j) > \bar{\nu}$ for all $k \neq 1, j$. It follows by Step 4 that the following equations hold:

$$\begin{aligned}
\frac{d\Pr(d=d_2|x^2)}{dx} \Big/ \frac{dc_{1,2}(x^2)}{dx} =& \varphi_1(\nu)\varphi_2(\nu)f(\nu) \\
\frac{d\Pr(d=d_3|x^3)}{dx} \Big/ \frac{dc_{1,3}(x^3)}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))\varphi_3(\nu)f(\nu) \\
\frac{d\Pr(d=d_4|x^5)}{dx} \Big/ \frac{dc_{1,4}(x^4)}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\varphi_4(\nu)f(\nu) \\
&\vdots \\
\frac{d\Pr(d=d_{d^*}|x^{d^*})}{dx} \Big/ \frac{dc_{1,d^*}(x^{d^*})}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\dots(1-\varphi_{d^*-1}(\nu))f(\nu)
\end{aligned}$$

The summation of these expressions recovers the quantity $\varphi_1(\nu)f(\nu)$. Next, substitute $\varphi_1(\nu)f(\nu)$ into the expressions above to sequentially recover $\{\varphi_j(\nu)\}_{2\leq j\leq d^*}$. Since ν can be made arbitrarily close to $\bar{\nu}$ by selecting a smaller value of ε and since $\alpha(\cdot)$ is continuous, $\lim_{\nu\to\bar{\nu}} \varphi_j(\nu) = \varphi_j(\bar{\nu}) = \varphi_j(1-\alpha(\bar{\nu})) = \varphi_j$ is also identified. Hence, $\{\varphi_j\}_{j=2}^{d^*}$ are identified.

Step 6: Identification of φ_1 and $\{\varphi_j\}_{d^*<j\leq D}$. The cutoffs are monotone in x^t and all cutoffs are on the right of $\bar{\nu}$ at $x^{\bar{\nu}}$. Consequently, $\Pr(d=d_j|x^{\bar{\nu}}) = 0$ for all $j > d^*$. Continuously decrease t until $\Pr(d=d_{j_1}|x^t) > 0$ for some $j_1 \in \mathcal{J} \equiv \{d^*+1,\dots,D\}$ and $\Pr(d=d_k|x^t) = 0$ for all $k \in \mathcal{J}\setminus\{j_1\}$. This will happen when $c_{d^*,j_1}(x^t)$ crosses $\bar{\nu}$, yielding

$$\frac{d\Pr(d=d_{j_1}|x^t)}{dx} \Big/ \frac{dc_{d*,j_1}(x^t)}{dx} = -\varphi_{j_1}(\nu)f(\nu)M_1(\nu),$$

where ν is in a small neighborhood near $\bar{\nu}$ s.t. $c_{k,j_1}(x^t) > \bar{\nu}$ for all $k > d^*, k \neq j_1$ and

$$M_1(\nu) \equiv \prod_{k\in\{2,\cdots,d^*-1\}: c_{k,j_1}(x^t)>\bar{\nu}} (1-\varphi_k(\nu)).$$

Near the end of this proof-step we will take the limit as $\nu \to \bar{\nu}$ after dividing out $f(\nu)$. Importantly, $M_1 \equiv \lim_{\nu\to\bar{\nu}} M_1(\nu)$ is known, since all relevant φ_k's are known, and M_1 does not depend on φ_1, since $c_{1,j_1}(x^t) < c_{d^*,j_1}(x^t) < \bar{\nu}$.

Next, continuously decrease t further until $\Pr(d=d_{j_2}|x^t) > 0$ for some $j_2 \in \mathcal{J}\setminus\{j_1\}$ and $\Pr(d=d_k|x^t) = 0$ for all $k \in \mathcal{J}\setminus\{j_1,j_2\}$. Again, this will happen when $c_{d^*,j_2}(x)$ crosses $\bar{\nu}$.

Hence,

$$\frac{d\,\mathrm{Pr}(d=d_{j_2}|x^t)}{dx} / \frac{dc_{d*,j_2}(x^t)}{dx} = -\varphi_{j_2}(\nu) f(\nu) M_2(\nu),$$
$$M_2(\nu) \equiv \prod_{k\in\{2,\ldots,d^*-1,j_1\}:c_{k,j_2}(x^t)>\bar{\nu}} (1-\varphi_k(\nu)).$$

The term $M_2 \equiv \lim_{\nu\to\bar{\nu}} M_2(\nu)$ is known, except possibly for the term $(1-\varphi_{j_1})$, since all other relevant φ_k's are known. The expression above defines $\varphi_{j_2}(\nu)$ as a strictly increasing function of $\varphi_{j_1}(\nu)$ regardless of whether $M_2(\nu)$ depends on $\varphi_{j_1}(\nu)$ or not. Indeed, for the case where $j_1 < j_2$ we have

$$\varphi_{j_2}(\nu) \propto \begin{cases} \varphi_{j_1}(\nu) & \text{if } c_{j_1,j_2}(x^t) < \bar{\nu} \\ \frac{\varphi_{j_1}(\nu)}{1-\varphi_{j_1}(\nu)} & \text{if } c_{j_1,j_2}(x^t) \geq \bar{\nu} \end{cases},$$

where the coefficients of proportionality are known in the limit. A similar expression holds when $j_1 > j_2$. This argument immediately extends to all $j \in \mathcal{J}$. In particular, for the case where $j_2 < j_3$

$$\varphi_{j_3} \propto \begin{cases} \varphi_{j_2(\nu)} & \text{if } c_{j_2,j_3}(x^t) < \bar{\nu} \\ \frac{\varphi_{j_2}(\nu)}{1-\varphi_{j_2}(\nu)} & \text{if } c_{j_2,j_3}(x^t) \geq \bar{\nu} \end{cases}.$$

Since $Pr(d = d_D|x^0) \neq 0$, the above sequential argument yields that $\varphi_D(\nu)$ is an increasing function of $\varphi_{j_1}(\nu)$. In turn, recall that $\varphi_1(\nu)f(\nu)$ is known for ν arbitrary close to $\bar{\nu}$. The limit of the ratio between $\varphi_1(\nu)f(\nu)$ and $\varphi_{j_1}(\nu)f(\nu)$, which is also known, yields φ_D as an increasing functions of φ_1. Hence, taken with the result in Step 3, the quantity φ_1 is uniquely pinned down. Identification of all other φ_j's immediately follow.

Step 7: Identification of $\alpha(\nu)$ and $f(\nu)$. The identification argument is iterative. For each alternative j, define

$$\Gamma_j^0 \equiv \{\nu \in [0,\bar{\nu}] : \exists x \in \mathcal{X} \text{ s.t. } \nu = c_{1,j}(x) \text{ and } c_j^*(x) \geq \bar{\nu}\}.$$

The set Γ_j^0 includes all preference parameters ν covered by the cutoff $c_{1,j}(\cdot)$ before any other relevant cutoffs for d_j enter the support. Let $\Gamma^0 \equiv \bigcap_{j=1}^D \Gamma_j^0$. By Step 4, Γ^0 is a non-trivial interval and $\bar{\nu} \in \Gamma^0$. For each $\nu \in \Gamma^0$ and each d_j, there is an $x^j \in \mathcal{X}$ such that $c_{1,j}(x^j) = \nu$. As a result, the following system of equations hold for each $\nu \in \Gamma^0$:

$$
\begin{aligned}
-\frac{d\Pr(d=d_2|x^2)}{dx}/\frac{dc_{1,2}(x^2)}{dx} =& \varphi_1(\nu)\varphi_2(\nu)f(\nu) \\
-\frac{d\Pr(d=d_3|x^3)}{dx}/\frac{dc_{1,3}(x^3)}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))\varphi_3(\nu)f(\nu) \\
-\frac{d\Pr(d=d_4|x^4)}{dx}/\frac{dc_{1,4}(x^4)}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\varphi_4(\nu)f(\nu) \\
& \vdots \\
-\frac{d\Pr(d=d_{d^*}|x^{d^*})}{dx}/\frac{dc_{1,d^*}(x^{d^*})}{dx} =& \varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\dots(1-\varphi_{d^*-1}(\nu))f(\nu),
\end{aligned}
\tag{A.4}
$$

The summation of these expressions recovers the quantity $\varphi_1(\nu)f(\nu)$. Substitute this into the first equation to obtain $\varphi_2(\nu) = \varphi_2(1-\alpha(\nu))$. But φ_2 is already known, so $\alpha(\nu)$ is identified on Γ^0. Finally, since $\varphi_1(\nu) = \varphi_1(1-\alpha(\nu))$ is now identified, so is $f(\nu)$ on Γ^0.

In the next step of the iteration, let $\bar{\nu}^1 = \min_{\nu\in\Gamma^0}\Gamma^0$ be the smallest value of ν where $\alpha(\nu)$ and $f(\nu)$ are identified. Define

$$
\Gamma_j^1 \equiv \{\nu\in[0,\bar{\nu}] : \exists x\in\mathcal{X} \text{ s.t. } \nu = c_{1,j}(x) \text{ and } c_j^*(x)\geq\bar{\nu}^1\} \quad \text{and} \quad \Gamma^1 \equiv \bigcap_{j=1}^{D}\Gamma_j^1.
$$

Then, a similar system to (A.4) holds $\forall\nu\in\Gamma^1$, but may include additional terms. These terms are known, because they are functions of $f(\cdot)$ and $\alpha(\cdot)$ evaluated at $\nu\in\Gamma^0$ (and also of $\{\varphi_j\}_{j=1}^D$). We can therefore repeat the argument from the base case to establish that $\alpha(\nu)$ and $f(\nu)$ are identified on Γ^1. We repeat this iterative procedure. After a finite number of steps T, we obtain $\Gamma^T = [0,\bar{\nu}]$ by Step 4; hence, $f(\cdot)$ and $\alpha(\cdot)$ are identified.

Edge Case I: Suppose $d^* = d_1$ (the identity is known from Step 2). The following expressions hold for ν close to $\bar{\nu}$:

$$
\begin{aligned}
-\frac{d\Pr(d=d_2|x^2)}{dx}/\frac{dc_{1,2}(x^2)}{dx} =& \varphi_2(\nu)f(\nu) \\
-\frac{d\Pr(d=d_3|x^3)}{dx}/\frac{dc_{1,3}(x^3)}{dx} =& (1-\varphi_2(\nu))\varphi_3(\nu)f(\nu) \\
-\frac{d\Pr(d=d_4|x^4)}{dx}/\frac{dc_{1,4}(x^4)}{dx} =& (1-\varphi_2(\nu))(1-\varphi_3(\nu))\varphi_4(\nu)f(\nu) \\
& \vdots \\
-\frac{d\Pr(d=d_D|x^D)}{dx}/\frac{dc_{1,D}(x^D)}{dx} =& (1-\varphi_2(\nu))(1-\varphi_3(\nu))\dots(1-\varphi_{D-1}(\nu))\varphi_D(\nu)f(\nu),
\end{aligned}
\tag{A.5}
$$

The ratio of the first and second equations in System (A.5) yields:

$$\frac{\varphi_2(\nu)}{\varphi_3(\nu)} = (1 - \varphi_2(\nu))\frac{1}{A_3(\boldsymbol{x}_3(\nu))},$$

where $\boldsymbol{x}(\nu) = (x^2(\nu), x^3(\nu), \dots, x^D(\nu))$ is a known implicit function of ν satisfying $\nu = c_{1,j}(x^j(\nu))$ for $j = 2, \dots, D$, and $A_3(\boldsymbol{x}(\nu))$ is data:

$$A_3(\boldsymbol{x}(\nu)) = \frac{\frac{d\Pr(d=d_3|x^3(\nu))}{dx} \Big/ \frac{dc_{1,3}(x^3(\nu))}{dx}}{\frac{d\Pr(d=d_2|x^2(\nu))}{dx} \Big/ \frac{dc_{1,2}(x^2(\nu))}{dx}}.$$

From this, we obtain

$$\begin{aligned}
\frac{\varphi_2}{\varphi_3} &= (1 - \varphi_2(1 - \alpha(\nu))\frac{1}{A_3(\boldsymbol{x}(\nu))} \\
\frac{A_3(\boldsymbol{x}(\nu))}{\varphi_3} &= \frac{1}{\varphi_2} - (1 - \alpha(\nu)) \\
\alpha(\nu) &= 1 - \frac{1}{\varphi_2} + \frac{A_3(\boldsymbol{x}(\nu))}{\varphi_3} \\
\alpha'(\nu) &= \left[\frac{\partial A_3(\boldsymbol{x}(\nu))}{\partial x^2}\frac{dx^2(\nu)}{d\nu} + \frac{\partial A_3(\boldsymbol{x}(\nu))}{\partial x^3}\frac{dx^3(\nu)}{d\nu}\right]\frac{1}{\varphi_3} \equiv B_3(\boldsymbol{x}(\nu))\frac{1}{\varphi_3},
\end{aligned}$$

where $B_3(\boldsymbol{x}(\nu))$ is a known function of data. A similar idea yields

$$\begin{aligned}
\alpha'(\nu) &= B_4(\boldsymbol{x}(\nu))\frac{1}{\varphi_4} \\
\alpha'(\nu) &= B_5(\boldsymbol{x}(\nu))\frac{1}{\varphi_5} \\
&\vdots \\
\alpha'(\nu) &= B_D(\boldsymbol{x}(\nu))\frac{1}{\varphi_D}.
\end{aligned}$$

Hence, the ratios $\frac{\varphi_3}{\varphi_4}, \frac{\varphi_4}{\varphi_5}, \frac{\varphi_5}{\varphi_6} \dots$ are identified. The ratio and the limit at the far end of the support of the third and fourth equations in System (A.5) yields

$$\begin{aligned}
0 &= 1 - \frac{1}{\varphi_3} + A_4(\boldsymbol{x}(\bar{\nu}))\frac{1}{\varphi_4} \\
\varphi_3 &= 1 - A_4(\boldsymbol{x}(\bar{\nu}))\frac{\varphi_3}{\varphi_4},
\end{aligned}$$

so φ_3 is identified and so are $\varphi_4, \dots, \varphi_D$. The ratio of the first and second equation in System (A.5) identifies φ_2. The proof continues on with Step 7. As such it does not require the second condition of the theorem.

Edge Case II: Suppose that $d^* = D$. From Step 5 we obtain φ_j for $j : 1 < j < D$. Next, we show how to identify φ_1. We can find x^j such that $\nu = c_{j,D}(x^j)$ is arbitrary close to zero satisfying $c_{j,k}(x^j) < 0$ for all $k \neq j, D$, and so the following system of equations holds:

$$\begin{aligned}
\frac{d\Pr(d = d_{D-1}|x^{D-1})}{dx} \Big/ \frac{dc_{D-1,D}(x^{D-1})}{dx} =& \varphi_{D-1}(\nu)f(\nu) \\
\frac{d\Pr(d = d_{D-2}|x^{D-2})}{dx} \Big/ \frac{dc_{D-2,D}(x^{D-2})}{dx} =& (1-\varphi_{D-1}(\nu))\varphi_{D-2}(\nu)f(\nu) \\
\frac{d\Pr(d = d_{D-3}|x^{D-3})}{dx} \Big/ \frac{dc_{D-3,D}(x^{D-3})}{dx} =& (1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\varphi_{D-3}(\nu)f(\nu) \\
&\vdots \\
\frac{d\Pr(d = d_1|x^1)}{dx} \Big/ \frac{dc_{1,D}(x^1)}{dx} =& (1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\dots(1-\varphi_2(\nu))\varphi_1(\nu)f(\nu).
\end{aligned}$$

The ratio of the first two equations yield

$$A = \frac{\varphi_{D-2}}{\varphi_{D-1}} \cdot (1 - \varphi_{D-1}(1 + \alpha(\nu)),$$

where A, φ_{D-2}, and φ_{D-1} are known terms; hence, $\alpha(0) = \lim_{\nu\to 0} \alpha(\nu)$ is identified. Once $\alpha(0)$ is identified, the term φ_1 is identified from the ratio of the first and last equations in the above system. Finally, $f(\nu)$ and $\alpha(\nu)$ are identified by Step 7.

Edge Case III: Suppose $d^* = 2$. By Steps 3, 5 and 6 all φ_j's are identified. A modified version of Steps 4 and 7 can be applied. We begin by starting at the lower end of the distribution with d_D taking the role of d_1; d_{D-1} taking the role of d_2, etc. Step 4 can be restated for

$$c_{j,D}(x) > c_j^{**}(x) \equiv \max_k \{\{c_{k,j}(x)\}_{1\le k<j}, \{c_{j,k}(x)\}_{j<k<D}\},\ \forall j.$$

Finally Step 7 can be repeated starting at the lower end of the support and building on the following equations

$$
\begin{aligned}
\frac{d\Pr(d=d_{D-1}|x^{D-1})}{dx} / \frac{dc_{D-1,D}(x^{D-1})}{dx} =& \varphi_D(\nu)\varphi_{D-1}(\nu)f(\nu) \\
\frac{d\Pr(d=d_{D-2}|x^{D-2})}{dx} / \frac{dc_{D-2,D}(x^{D-2})}{dx} =& \varphi_D(\nu)(1-\varphi_{D-1}(\nu))\varphi_{D-2}(\nu)f(\nu) \\
\frac{d\Pr(d=d_{D-3}|x^{D-3})}{dx} / \frac{dc_{D-3,D}(x^{D-3})}{dx} =& \varphi_D(\nu)(1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\varphi_{D-3}(\nu)f(\nu) \\
&\vdots \\
\frac{d\Pr(d=d_3|x^3)}{dx} / \frac{dc_{1,3}(x^3)}{dx} =& \varphi_D(\nu)(1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\ldots(1-\varphi_3(\nu))f(\nu).
\end{aligned}
$$

□

Proof of Theorem 7. Let ν, x, $\mathcal{N}_\epsilon(x) = \{x' : \|x' - x\| < \epsilon\}$ satisfy the conditions in the theorem. Then $\nu = c_{j,k}(x)$ for all j, k. For any pair of alternatives (d_j, d_k) we can perturb x_k, $k \notin \{j, d^*\}$, and x_l, $\forall l \notin \{j, d^*, k\}$, so that the resulting $x' \in \mathcal{N}_\epsilon(x)$ is such that

$$
\begin{aligned}
&U_\nu(d_k, x'_k) > U_\nu(d_j, x_j) \\
&U_\nu(d_j, x_j) > U_\nu(d_l, x'_l), \quad \forall l \in \mathcal{D} \setminus \{j, k, d^*\}.
\end{aligned}
$$

And we can do another perturbation of x_l, $\forall l \notin \{j, d^*\}$, so that the resulting $x'' \in \mathcal{N}_\epsilon(\tilde{x})$ is such that

$$
U_\nu(d_j, x_j) > U_\nu(d_l, x''_l), \quad \forall l \in \mathcal{D} \setminus \{j, d^*\}.
$$

Then

$$
\begin{aligned}
\frac{\partial \Pr(d=d_j|x')}{\partial x_{d^*}} &= \varphi_j(x_j,\nu)(1-\varphi_k(x'_k,\nu))f(\nu)\frac{\partial c_{j,d^*}(x)}{\partial x_{d^*}} \\
\frac{\partial \Pr(d=d_j|x'')}{\partial x_{d^*}} &= \varphi_j(x_j,\nu)f(\nu)\frac{\partial c_{j,d^*}(x)}{\partial x_{d^*}}.
\end{aligned}
$$

Taking the ratio of the expressions above identifies $(1-\varphi_k(x'_k,\nu))$. By continuity we identify $\varphi_k(x_k,\nu)$. Identical steps identify $\varphi_j(x_j,\nu)$, $\forall j \neq d^*$, and hence $f(\nu)$.

□

Proof of Proposition 4. Take any non-empty consideration set $\mathcal{K}$. For a given preference coefficient ν, let $j_{\mathcal{K}}(x,\nu)$ denote the identity of the best alternative in this consideration set. By the natural ordering, $j_{\mathcal{K}}(x,\nu)$ is an increasing step function in ν. Hence, $I(j_{\mathcal{K}}(x,\nu) \leq J)$ is a decreasing step function. The term $\Pr\left(\bigcup_{j=1}^{J} d_j \middle| x, \nu\right)$ is a non-negatively weighted sum of $I(j_{\mathcal{K}}(x,\nu) \leq J)$. Hence it is decreasing in ν. □

Proof of Proposition 5. Consider a limited consideration model with preferences $U_\nu(d_j, x)$ and consideration probability $\mathcal{Q}_\nu^x(\mathcal{K})$, $\mathcal{K} \subset \mathcal{D}$. The optimal choice from $\mathcal{D}$ conditional on the DM facing the consideration set $\mathcal{K} \neq \emptyset$ is the alternative with the largest value of $U_\nu(d_j, x)$ subject to $j \in \mathcal{K}$. This is the same solution as the one that maximizes $V_\nu(d_j, x, \epsilon_j)$ where $\epsilon_j = 0$ for all $j \in \mathcal{K}$ and $\epsilon_j = -\infty$ for all $j \in \mathcal{D} \setminus \mathcal{K}$. Finally, since conditional on x the consideration set $\mathcal{K}$ has the same distribution as the set of alternatives with $\epsilon_j = 0$ (this is by construction), the limited consideration model and this ORUM model yield the same model predictions, and hence they are equivalent. □

B Application: Verifying Cutoff Order

We start by recalling that CARA and CRRA utility functions satisfy the following basic property (see, e.g., Pratt, 1964; Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2018).[1]

Lemma B.1. *For any $y_0 > y_1 > y_2 > 0$, the ratio $R(y_0, y_1, y_2) \equiv \frac{u_\nu(y_1) - u_\nu(y_2)}{u_\nu(y_0) - u_\nu(y_1)}$ is strictly increasing in ν.*

It follows that CARA and CRRA utility functions also satisfy a slightly extended version of the property above:

Lemma B.2. *For any $y_0 > y_1 > y_2 > y_3 > 0$, the ratio $M_\nu(y_0, y_1, y_2, y_3) \equiv \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)}$ is strictly increasing in ν.*

Proof.

$$\begin{aligned} M_\nu(y_0, y_1, y_2, y_3) &= \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)} = \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_1) - u_\nu(y_2)} \times \frac{u_\nu(y_1) - u_\nu(y_2)}{u_\nu(y_0) - u_\nu(y_1)} \\ &= R_\nu(y_1, y_2, y_3) R_\nu(y_0, y_1, y_2) \end{aligned}$$

□

[1] This property is equivalent to condition (e) in Pratt (1964, Theorem 1). As shown there, it is equivalent to assuming that an increase in ν corresponds to an increase in the coefficient of absolute risk aversion.

For our application, we show that $c_{1,j}(\bar{p},\mu) < c_{1,j+1}(\bar{p},\mu)$ for any $j \geq 2$ under both CARA and CRRA preferences.

Proposition B.1. *Suppose deductibles and prices are such that*

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} < \frac{d_1 - d_j}{d_1 - d_{j+1}}$$

and $d_k + p_k < w$ for all k. Under either CARA or CRRA expected utility preferences, the cutoff mapping is unique and satisfies $c_{1,j}(\bar{p},\mu) < c_{1,j+1}(\bar{p},\mu)$ for all $j > 1$.

Note that in a perfectly competitive markets where additional coverage is simply proportional to its price both ratios will be equal. In practice, however, one might expect that with some market power the prices increase faster than then coverage, and hence

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} < \frac{d_1 - d_j}{d_1 - d_{j+1}}$$

This is exactly what we find in our data (as well as for a larger number of firms appearing in Barseghyan, Prince, & Teitelbaum (2011)).

Proof. We start with CARA preferences. The existence and the uniqueness of $c_{j,k}(x)$ for all $j < k$ follows directly from the Lemma B.2. Indeed note that $p_j < p_k < p_k + d_k < p_j + d_j$.[2] At the cutoff the DM is indifferent between lotteries j and k. Equating two expected utilities and rearranging we have that

$$\frac{e^{-\nu(w-p_k-d_k)} - e^{-\nu(w-p_j-d_j)}}{e^{-\nu(w-p_j)} - e^{-\nu(w-p_k)}} = \frac{1-\mu}{\mu}, \tag{B.1}$$

where w is the DM's initial wealth. By Lemma B.2, the L.H.S. of Equation B.1 is strictly monotone in ν, and it tends to $+\infty$ when ν goes to $+\infty$ and to zero when ν goes to $-\infty$. It follows that there exists a unique ν, i.e the cutoff $c_{j,k}(x)$, that solves the Equation B.1. Moreover, since the L.H.S. is strictly monotone in ν it follows from the Implicit Function Theorem that $c_{j,k}(x)$ is continuous in μ and $\bar{p}$.

The next step is to establish $c_{1,j}(\bar{p},\mu) < c_{1,j+1}(\bar{p},\mu)$, $j > 1$. For the purpose of obtaining a contradiction, suppose that there exists $(\bar{p},\mu)$ and an j such that $c_{1,j}(\bar{p},\mu) \geq c_{1,j+1}(\bar{p},\mu)$.

[2] If $p_k + d_k > p_j + d_j$, then alterantive j first order stochastically dominates k and hence the cuttoff is $+\infty$.

Since the expected utility of lottery k is proportional to

$$EU_\nu(L_k) \propto -e^{\nu p_k}\left(1-\mu+\mu e^{\nu d_k}\right),$$

there exists $\nu = c_{1,j}(\bar{p},\mu) \geq c_{1,j+1}(\bar{p},\mu)$ such that

$$\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_j}} e^{\nu(g_1-g_j)\bar{p}} = 1 \leq \frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_{j+1}}} e^{\nu(g_1-g_{j+1})\bar{p}}$$

Taking logs yields

$$\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_j}}\right) = -\nu(g_1-g_j)\bar{p}$$
$$\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_{j+1}}}\right) \geq -\nu(g_1-g_{j+1})\bar{p}.$$

Dividing through and using the fact that $-\nu(g_1-g_{j+1})\bar{p} \geq -\nu(g_1-g_j)\bar{p} \geq 0$ yields

$$\frac{\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_j}}\right)}{\log\left(\frac{1-\mu+\mu e^{\nu d_1}}{1-\mu+\mu e^{\nu d_{j+1}}}\right)} \leq \frac{g_1-g_j}{g_1-g_{j+1}}.$$

The R.H.S. is less than one. We claim that the L.H.S. is monotonically decreasing in $\mu<1$. To show this, denote $\hat{\mu} = \frac{1-\mu}{\mu}$, $\Delta_1 = e^{\nu d_1}$, $\Delta_j = e^{\nu d_j}$, and $\Delta_{j+1} = e^{\nu d_{j+1}}$ to rewrite the L.H.S. as follows

$$\text{L.H.S} = f\left(\frac{1-\mu}{\mu}\right) = f(\hat{\mu}) \equiv \frac{\log(\Delta_1+\hat{\mu}) - \log(\Delta_j+\hat{\mu})}{\log(\Delta_1+\hat{\mu}) - \log(\Delta_{j+1}+\hat{\mu})}.$$

First, we show that the above expression is monotonically increasing in $\hat{\mu}$. Observe that

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \left(\frac{1}{\Delta_1+\hat{\mu}} - \frac{1}{\Delta_j+\hat{\mu}}\right)\frac{1}{\log(\Delta_1+\hat{\mu})-\log(\Delta_j+\hat{\mu})} - \left(\frac{1}{\Delta_1+\hat{\mu}} - \frac{1}{\Delta_{j+1}+\hat{\mu}}\right)\frac{1}{\log(\Delta_1+\hat{\mu})-\log(\Delta_{j+1}+\hat{\mu})}$$

After relabeling $\Lambda_1 = -\log(\Delta_1+\hat{\mu})$, $\Lambda_j = -\log(\Delta_j+\hat{\mu})$ and $\Lambda_{j+1} = -\log(\Delta_{j+1}+\hat{\mu})$ we obtain

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \frac{e^{\Lambda_1}-e^{\Lambda_{j+1}}}{\Lambda_1-\Lambda_{j+1}} - \frac{e^{\Lambda_1}-e^{\Lambda_j}}{\Lambda_1-\Lambda_j}.$$

Since $\Lambda_1 < \Lambda_j < \Lambda_{j+1}$ and exponential function is convex, the expression above is positive. Thus, the derivative of $f\left(\frac{1-\mu}{\mu}\right)$ W.R.T. μ is negative as claimed. That is, $f\left(\frac{1-\mu}{\mu}\right)$ achieves its lowest value at $\mu = 1$ and is equal to $\frac{d_1 - d_j}{d_1 - d_{j+1}}$. Finally, a contradiction is obtained since

$$\text{L.H.S} \geq \min_{\mu} f\left(\frac{1-\mu}{\mu}\right) = \frac{d_1 - d_j}{d_1 - d_{j+1}} > \frac{g_1 - g_j}{g_1 - g_{j+1}} = \text{R.H.S},$$

where the strict inequality is by assumption. Therefore, $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$ under CARA as claimed.

Under CRRA, $c_{j,k}(\bar{p}, \mu)$ exist and are continuous exactly for the same reasons as under CARA. It remains to establish that $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$. For the purpose of obtaining a contradiction, suppose $c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ for some $(\bar{p}, \mu)$. Consider the following Taylor expansion for the CRRA Bernoulli utility function $u_\nu(w) \equiv \frac{w^{1-\nu}}{1-\nu}$:

$$\frac{(w - p_k)^{1-\nu}}{1-\nu} = \frac{w^{1-\nu}}{1-\nu} + \frac{w^{-\nu}}{1!}(-p_k) - \nu\frac{w^{-\nu-1}}{2!}(-p_k)^2 + \nu(\nu+1)\frac{w^{-\nu-2}}{3!}(-p_k)^3 + \ldots$$

Or, equivalently,

$$(1-\nu)w^{\nu-1}[u_\nu(w - p_k) - u_\nu(w)] = (\nu - 1)\frac{1}{1!}w^{-1}p_k + (\nu-1)\nu\frac{w^{-2}}{2!}p_k^2 + (\nu-1)\nu(\nu+1)\frac{w^{-3}}{3!}p_k^3 + \ldots$$

Hence,

$$EU_\nu(L_k) \propto (1-\mu)\sum_{t=1}^{\infty}\omega_t(\nu)p_k^t + \mu\sum_{t=1}^{\infty}\omega_t(\nu)\left(p_k + d_k\right)^t.$$

where $\omega_t(\nu) \equiv (t!w^t)^{-1}\prod_{t'=0}^{t-1}(\nu - 1 + t') < 0$ when $\nu \in (0,1)$. When $\nu > 1$, $\omega_t(\nu) > 0$ but the factor premultiplying $u_\nu(w - p_k)$ above is negative, so we would still come to the same conclusion that $EU_\nu(L_k)$ is proportional to a power series with coefficients $\tau_t(\nu) = -\omega_t(\nu) < 0$. The power series are absolutely convergence provided that $p_k + d_k < w$, so the difference in the power series for $EU_\nu(L_j)$ and $EU_\nu(L_k)$ is equal to the sum of the difference:

$$\begin{aligned} EU_\nu(L_j) - EU_\nu(L_k) &\propto (1-\mu)\sum_{t=1}^{\infty}\omega_t(\nu)\left(p_j^t - p_k^t\right) + \mu\sum_{t=1}^{\infty}\omega_t(\nu)\left((p_j + d_j)^t - (p_k + d_k)^t\right) \\ &= (p_j - p_k)(1-\mu)\sum_{t=1}^{\infty}\omega_t(\nu)\sum_{h=0}^{t-1}p_j^h p_k^{t-h} + \\ &+ ((p_j - p_k) + (d_j - d_k))\,\mu\sum_{t=1}^{\infty}\omega_t(\nu)\sum_{h=0}^{t-1}(p_j + d_j)^h (p_k + d_k)^{t-h}. \end{aligned}$$

The condition $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ implies

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} \geq \frac{p_1 - p_j + d_1 - d_j}{p_1 - p_{j+1} + d_1 - d_{j+1}} \delta(\nu), \tag{B.2}$$

where

$$\delta(\nu) \equiv \frac{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} (p_1 + d_1)^h (p_j + d_j)^{t-h}}{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} (p_1 + d_1)^h (p_{j+1} + d_{j+1})^{t-h}} \frac{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} p_1^h p_{j+1}^{t-h}}{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} p_1^h p_j^{t-h}}.$$

Under the assumption $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ it is also the case that $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu) \geq c_{j,j+1}(\bar{p}, \mu)$ by Fact 4. Hence, $p_j + d_j > p_{j+1} + d_{j+1}$. Indeed otherwise $p_{j+1} - p_j > d_j - d_{j+1}$ is a violation of the first order stochastic dominance. Taken with $p_{j+1} > p_j$, it follows that $\delta(\nu) > 1$. Finally, a contradiction will be obtained if

$$\frac{p_j - p_1}{p_{j+1} - p_1} \leq \frac{p_{j+1} - p_1 + d_1 - d_j}{p_1 - p_{j+1} + d_1 - d_{j+1}},$$

since then Equation (B.2) will not hold. Re-arranging this expression we obtain:

$$\begin{aligned} \frac{p_1 - p_{j+1} + d_1 - d_{j+1}}{p_{j+1} - p_1} &\leq \frac{p_1 - p_j + d_1 - d_j}{p_j - p_1} \\ \frac{d_1 - d_{j+1}}{p_{j+1} - p_1} &\leq \frac{d_1 - d_j}{p_j - p_1} \\ \frac{p_1 - p_j}{p_1 - p_{j+1}} &\leq \frac{d_1 - d_j}{d_1 - d_{j+1}}. \end{aligned}$$

The latter inequality holds by assumption. It follows that $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$, $j > 1$.

□

C Monetary Cost of Limited Consideration

We view limited consideration as a mechanism that constrains households from achieving their first-best alternative either because the market setting forces some alternatives to become more salient than others (e.g. agent effects) or because of time or psychological costs that prevent the household from evaluating all alternatives in the choice set. Regardless of the underlying mechanism(s) of limited consideration, we can quantify its *monetary* cost within our framework. We ask, *ceteris paribus*, how much money the households "leave on the table" when choosing deductibles in property insurance under limited consideration rather than under full consideration. This is likely to be a lower bound on actual monetary losses arising from limited consideration, because insurance companies might be exploiting sub-optimality of households choices when setting prices or choosing menus.

We measure the monetary costs of limited consideration as follows. For each household we compute (the expected value of) the certainty equivalent of the lottery associated with the households' optimal choice, as well as of the one associated with their choice under limited consideration.[3] We then take the difference between these certainty equivalent values and average them across all households in the sample. On average, we find that households lose \$50 dollars across the three deductibles because of limited consideration. See Table E.6 for variation conditional on demographic characteristics and insurance score. We also find wide dispersion in loss across households (see Figure E.6). In particular, the 10^{th} percentile of losses is \$31 and the 90^{th} is \$73.

[3]Certainty equivalent of the lottery is defined as the minimum amount they are willing to accept in lieu of the lottery. In our case, for alternative j, it is simply $ce_j \equiv \frac{1}{\nu}\ln[(1-\mu)\exp(\nu p_j) + \mu\exp(\nu(p_j + d_j))]$.

D Data

Table D.1 Descriptive Statistics

Variable	Mean	Std. Dev.	1st %	99th %
Age	53.9	15.6	25.4	84.1
Female	0.40			
Single	0.22			
Married	0.55			
Second Driver	0.43			
Insurance Score	767	112	532	985

Table D.2 Frequency of Deductible Choices Across Contexts

Deductible	1000	500	250	200	100	50
Collision	0.063	0.676	0.122	0.129	0.009	
Comprehensive	0.037	0.430	0.121	0.329	0.039	0.044
Home	0.176	0.559	0.262		0.002	

Table D.3 Deductible Rank Correlations Across Contexts

	Collision	Comprehensive	Home
Collision	1		
Comprehensive	0.61	1	
Home	0.37	0.35	1

Table D.4 Average Premiums Across Coverages

Deductible	1,000	500	250	200	100	50
Collision	145	187	243	285	327	
Comprehensive	94	117	147	155	178	224
Home	594	666	720		885	

E Empirical Results: Figures and Tables

E.1 The ARC Model with Observable Demographics

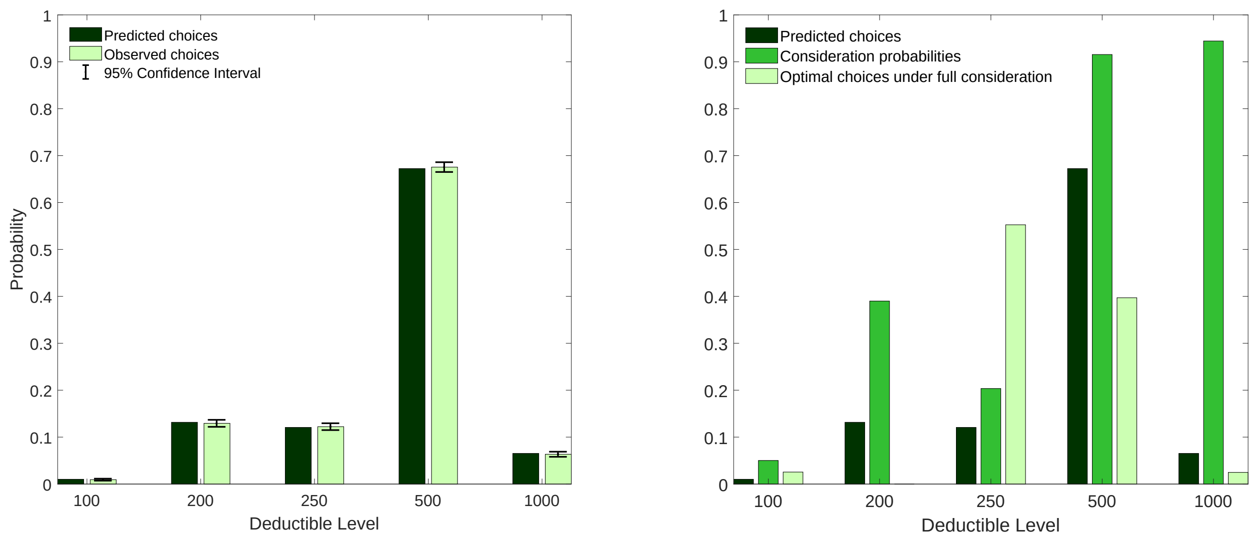
While it is ideal to control for households' observable characteristics non-parametrically, it is data demanding. In practice, it is commonly assumed that household characteristics shift the expected value of the preference-coefficient distribution.[4] We adopt the same strategy here by assuming that for each household i, $\log \frac{\beta_{1,i}}{\beta_2} = \mathbf{Z}_i\gamma$, where γ is an unknown vector to be estimated. The terms $\beta_{1,i}$ and β_2 denote the parameters of the Beta distribution, where $\beta_{1,i}$ is household specific and β_2 is common across households. The preference coefficients are random draws from a distribution with an expected value that is a function of the observable characteristics given by $E(\nu_i) = \frac{\beta_{1,i}}{\beta_{1,i}+\beta_2}\bar{\nu} = \frac{e^{\mathbf{Z}_i\gamma}}{1+e^{\mathbf{Z}_i\gamma}}\bar{\nu}$.[5] The results of this estimation are in line with our first estimation. (See Column 2 in Table E.1, as well as Figures E.1 and E.2.) The new observation here is that the model closely matches the distribution of choices across various sub-populations in the sample including gender, age, credit worthiness, and contracts with multiple drivers. The model's ability to match these conditional distributions can be attributed, in part, to the dependence of risk preferences on household characteristics. The model is, however, fairly parsimonious as the consideration parameters are restricted to be the same across all households. Finally, estimated consideration probabilities are close in magnitude to those estimated above. In particular, the highest deductibles ($\$1,000$ and $\$500$) are most likely to be considered, with respective frequencies of 0.94 and 0.92. The remaining alternatives are considered at much lower frequencies.

[4]For exmaple, Cohen & Einav (2007) assume that $\log \nu_i = \mathbf{Z}_i\gamma + \varepsilon_i$, where $\mathbf{Z}_i$ are the observables for household i and ε_i is i.i.d. $N(0, \sigma^2)$. Hence, $E(\nu_i) = e^{\mathbf{Z}_i\gamma+\sigma^2/2}$.

[5]If, instead, we assume $\log \frac{\beta_{2,i}}{\beta_1} = \mathbf{Z}_i\tilde{\gamma}$, then we arrive to the same expression for the expected value with the exception that $\tilde{\gamma} = -\gamma$.

E.2 Figures

Figure E.1: The ARC Model with Observable Demographics



The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

Figure E.2: The ARC Model with Observable Demographics: Conditional Distributions

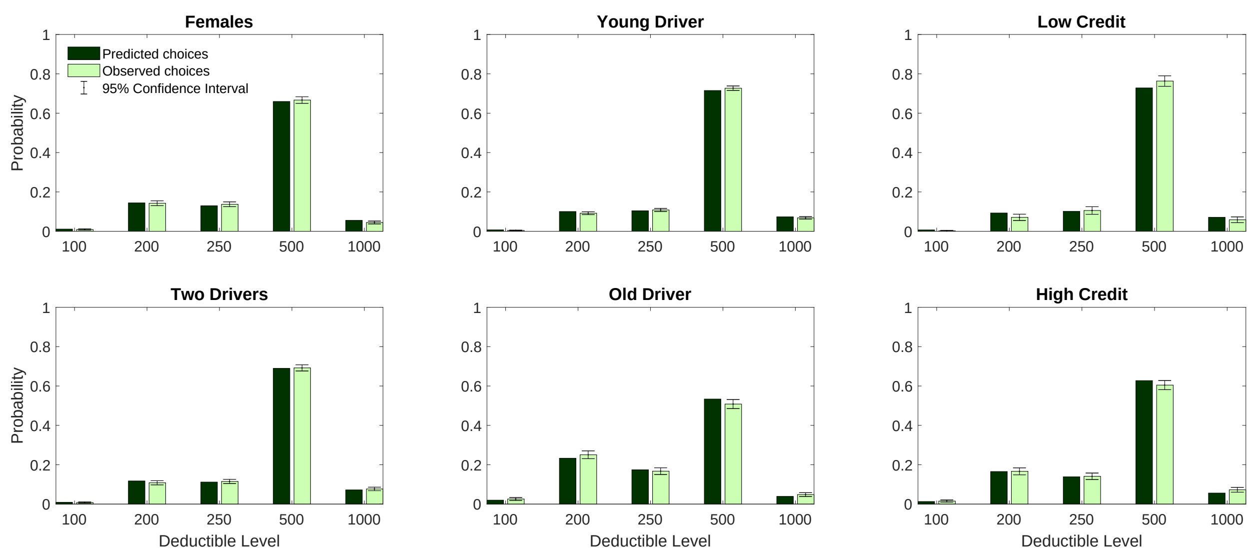


Figure E.3: The Mixed Logit

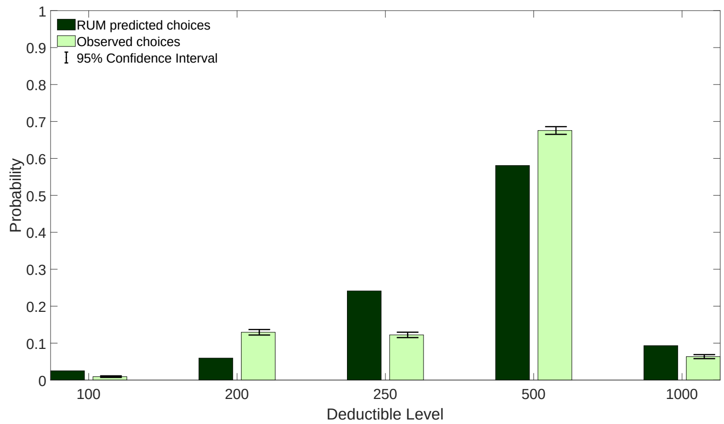
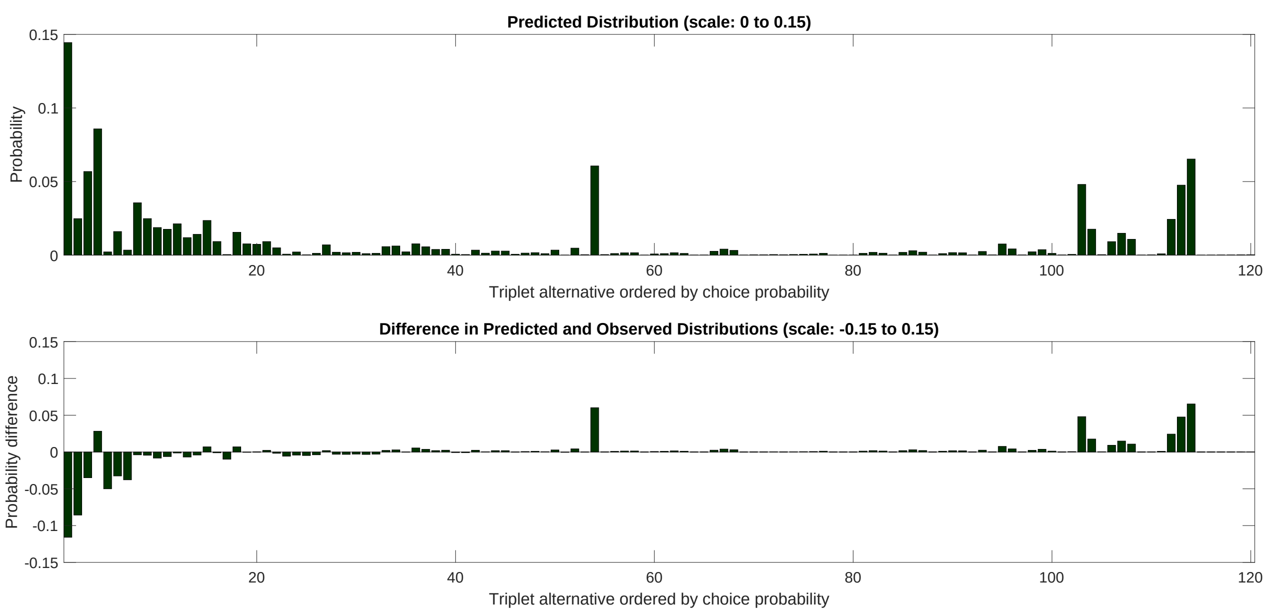
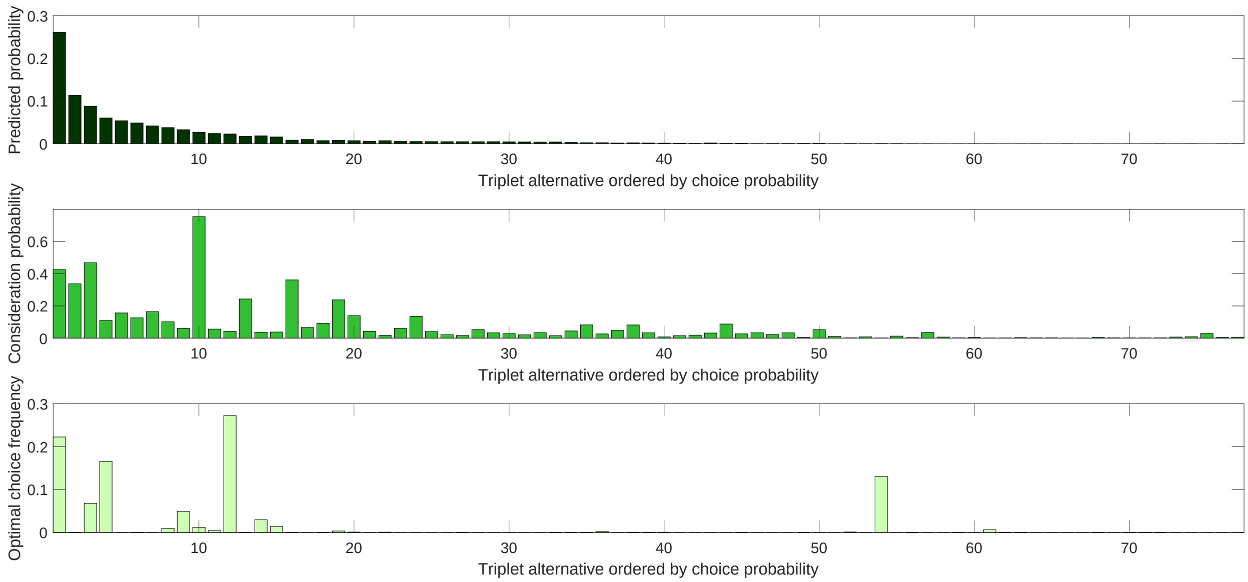


Figure E.4: The Mixed Logit, Three Coverages



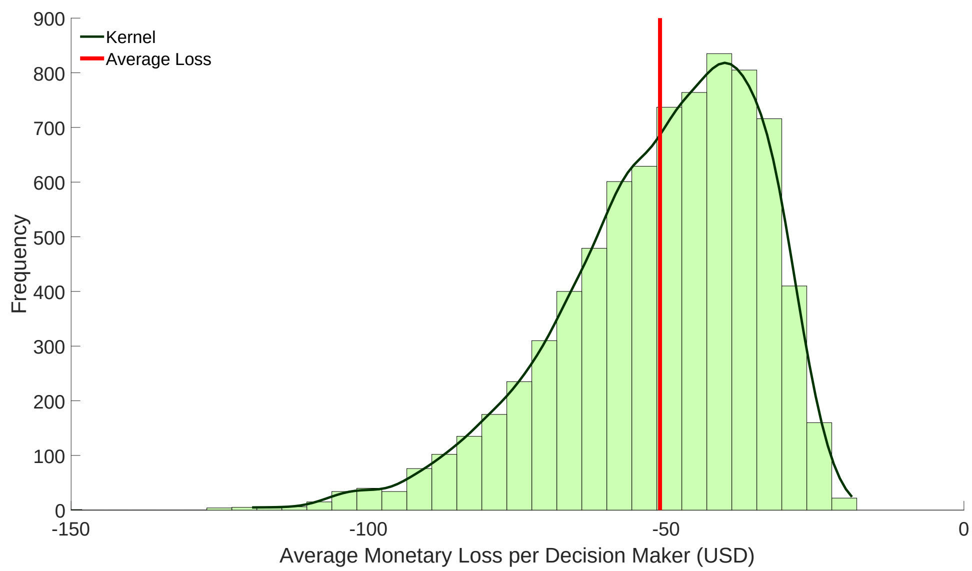
Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

Figure E.5: The ARC Model, Three Coverages:
Consideration and Optimal Choice Distribution



Triplets are sorted by observed frequency at which they are chosen.

Figure E.6: The ARC Model with Three Coverages:
Monetary Loss From Limited Consideration



E.3 Tables

Table E.1 MLE Estimation Results for the ARC Model: Collision Only

	ARC Model		**ARC Model with Observables**	
Average β_{1i}	1.70	[1.56, 1.82]	2.23	[1.93, 2.50]
β_2	7.45	[6.68, 8.08]	9.20	[8.09, 10.1]
Mean of ν	0.0037	[0.0036, 0.0038]	0.0038	[0.0036, 0.0040]
SD of ν	0.0024	[0.0023, 0.0026]	0.0022	[0.0021, 0.0023]
Intercept	-	-	-1.41	[-1.47, -1.33]
Age	-	-	0.207	[0.173, 0.237]
Age^2	-	-	0.048	[0.022, 0.073]
Female Driver	-	-	0.077	[0.051, 0.104]
Single Driver	-	-	0.050	[0.022, 0.079]
Married Driver	-	-	0.103	[0.062, 0.144]
Credit Score	-	-	0.134	[0.107, 0.160]
2+ Drivers	-	-	-0.302	[-0.370, -0.224]
Collision $100	0.059	[0.050, 0.068]	0.050	[0.042, 0.058]
Collision $200	0.412	[0.391, 0.433]	0.390	[0.364, 0.413]
Collision $250	0.206	[0.198, 0.214]	0.204	[0.193, 0.212]
Collision $500	0.920	[0.913, 0.926]	0.915	[0.909, 0.924]
Collision $1000	1.000	[1.000, 1.000]	0.944	[0.899, 1.000]

Table E.2 MLE Estimation Results for the Proportionally Shifting Consideration Model

	ARC Model		**ARC Model with Observables**	
Average β_{1i}	1.44	[1.31, 1.55]	2.11	[1.86, 2.28]
β_2	6.07	[5.35, 6.67]	8.74	[7.73, 9.58]
Mean of ν	0.0038	[0.0037, 0.0040]	0.0038	[0.0036, 0.0040]
SD of ν	0.0027	[0.0026, 0.0028]	0.0023	[0.0021, 0.0024]
Intercept	-	-	-1.40	[-1.47, -1.35]
Age	-	-	0.194	[0.160, 0.222]
Age^2	-	-	0.036	[0.010, 0.059]
Female Driver	-	-	0.070	[0.046, 0.096]
Single Driver	-	-	0.049	[0.021, 0.076]
Married Driver	-	-	0.091	[0.047, 0.130]
Credit Score	-	-	0.135	[0.110, 0.160]
2+ Drivers	-	-	-0.283	[-0.348, -0.200]
Collision $100	0.061	[0.051, 0.070]	0.055	[0.046, 0.063]
Collision $200	0.424	[0.401, 0.446]	0.408	[0.382, 0.433]
Collision $250	0.211	[0.202, 0.220]	0.212	[0.201, 0.222]
Collision $500	0.985	[0.974, 0.998]	0.961	[0.929, 0.977]
Collision $1000	1.000	-	1.000	-
Average ξ_{1i}	0.478	[0.277, 0.652]	0.148	[0.021, 0.212]
ξ_2	26.7	[16.1, 37.9]	7.14	[0.939, 10.3]
ξ_1**: Intercept**	-	-	-2.24	[-3.59, 0.0011]
ξ_1**: Age**	-	-	1.24	[-0.736, 1.75]
ξ_1**: Age**^{2}	-	-	-0.382	[-0.701, 0.584]
ξ_1**: Female Driver**	-	-	-0.323	[-1.16, 0.910]
ξ_1**: Single Driver**	-	-	0.382	[-1.51, 0.650]
ξ_1**: Married Driver**	-	-	0.0017	[-2.35, 1.45]
ξ_1**: Credit Score**	-	-	0.405	[-0.688, 0.642]
ξ_1**: 2+ Drivers**	-	-	0.485	[-2.22, 1.98]

Table E.3 MLE Estimation Results for the Mixed Logit:
Collision Only

	Mixed Logit	
Average β_{1i}	9.07	[7.54, 10.2]
β_2	124.4	[106.0, 137.5]
Mean of ν	0.0014	[0.0013, 0.0014]
SD of ν	0.0004	[0.0004, 0.0005]
Intercept	-2.59	[-2.63, -2.55]
Age	-0.139	[-0.156, -0.122]
Age^2	-0.024	[-0.037, -0.010]
Female Driver	-0.0035	[-0.019, 0.012]
Single Driver	-0.0098	[-0.026, 0.0060]
Married Driver	-0.030	[-0.054, -0.0078]
Credit Score	0.091	[0.076, 0.105]
2+ Drivers	-0.016	[-0.061, 0.029]
Sigma	0.039	[0.037, 0.041]

Table E.4 MLE Estimation Results for the ARC Model, Three Coverages

	ARC Model			ARC Model	
Average β_{1i}	4.70	[3.89, 5.30]	**(250,200,250)**	0.037	[0.033, 0.041]
β_2	24.0	[19.7, 27.2]	**(250,200,500)**	0.056	[0.051, 0.061]
Mean of ν	0.0032	[0.0032, 0.0033]	**(250,200,1000)**	0.045	[0.035, 0.055]
SD of ν	0.0013	[0.0012, 0.0014]	**(250,250,100)**	0.0011	[0.0004, 0.0020]
Intercept	-1.68	[-1.72, -1.64]	**(250,250,250)**	0.042	[0.038, 0.046]
Age	0.162	[0.146, 0.180]	**(250,250,500)**	0.061	[0.056, 0.066]
Age^2	0.041	[0.026, 0.054]	**(250,250,1000)**	0.026	[0.019, 0.034]
Female Driver	0.043	[0.027, 0.061]	**(250,500,500)**	0.0007	[0.0004, 0.0013]
Single Driver	0.010	[-0.0075, 0.028]	**(500,50,250)**	0.034	[0.028, 0.042]
Married Driver	0.030	[0.0054, 0.054]	**(500,50,500)**	0.053	[0.044, 0.063]
Credit Score	0.137	[0.121, 0.153]	**(500,50,1000)**	0.033	[0.018, 0.047]
2+ Drivers	-0.097	[-0.141, -0.052]	**(500,100,250)**	0.015	[0.012, 0.019]
(100,50,250)	0.041	[0.033, 0.049]	**(500,100,500)**	0.043	[0.036, 0.049]
(100,50,500)	0.015	[0.0099, 0.021]	**(500,100,1000)**	0.048	[0.034, 0.062]
(100,50,1000)	0.013	[0.0048, 0.023]	**(500,200,100)**	0.0079	[0.0040, 0.012]
(100,100,100)	0.0023	[0.0009, 0.0042]	**(500,200,250)**	0.126	[0.119, 0.133]
(100,100,250)	0.0077	[0.0049, 0.010]	**(500,200,500)**	0.337	[0.322, 0.352]
(100,100,500)	0.0050	[0.0027, 0.0078]	**(500,200,1000)**	0.243	[0.221, 0.264]
(100,100,1000)	0.0047	[0.0019, 0.0086]	**(500,250,100)**	0.0018	[0.0008, 0.0032]
(100,200,250)	0.0005	[0.0002, 0.0010]	**(500,250,250)**	0.038	[0.034, 0.042]
(100,200,500)	0.0008	[0.0004, 0.0015]	**(500,250,500)**	0.102	[0.095, 0.109]
(100,200,1000)	0.0036	[0.0013, 0.0065]	**(500,250,1000)**	0.093	[0.080, 0.106]
(200,50,100)	0.011	[0.0052, 0.016]	**(500,500,100)**	0.0032	[0.0015, 0.0059]
(200,50,250)	0.066	[0.057, 0.075]	**(500,500,250)**	0.110	[0.103, 0.116]
(200,50,500)	0.061	[0.051, 0.071]	**(500,500,500)**	0.426	[0.413, 0.439]
(200,50,1000)	0.033	[0.018, 0.048]	**(500,500,1000)**	0.469	[0.450, 0.486]
(200,100,100)	0.0020	[0.0008, 0.0037]	**(1000,50,250)**	0.0069	[0.0020, 0.013]
(200,100,250)	0.021	[0.017, 0.026]	**(1000,50,500)**	0.0085	[0.0022, 0.015]
(200,100,500)	0.028	[0.023, 0.034]	**(1000,50,1000)**	0.029	[0.0031, 0.053]
(200,100,1000)	0.023	[0.012, 0.033]	**(1000,100,250)**	0.0049	[0.0019, 0.0090]
(200,200,100)	0.0017	[0.0007, 0.0032]	**(1000,100,500)**	0.0060	[0.0021, 0.011]
(200,200,250)	0.157	[0.147, 0.167]	**(1000,100,1000)**	0.035	[0.0080, 0.064]
(200,200,500)	0.165	[0.154, 0.176]	**(1000,200,250)**	0.032	[0.019, 0.045]
(200,200,1000)	0.136	[0.113, 0.158]	**(1000,200,500)**	0.082	[0.061, 0.105]
(200,250,250)	0.0004	[0.0002, 0.0006]	**(1000,200,1000)**	0.088	[0.046, 0.128]
(200,250,500)	0.0005	[0.0002, 0.0009]	**(1000,250,250)**	0.0067	[0.0027, 0.012]
(200,500,250)	0.0015	[0.0008, 0.0023]	**(1000,250,500)**	0.027	[0.015, 0.039]
(200,1000,1000)	0.0047	[0.0017, 0.0085]	**(1000,250,1000)**	0.053	[0.025, 0.083]
(250,50,100)	0.0020	[0.0008, 0.0037]	**(1000,500,250)**	0.033	[0.022, 0.044]
(250,50,250)	0.021	[0.017, 0.025]	**(1000,500,500)**	0.140	[0.119, 0.161]
(250,50,500)	0.033	[0.027, 0.039]	**(1000,500,1000)**	0.362	[0.309, 0.405]
(250,100,250)	0.017	[0.015, 0.020]	**(1000,1000,250)**	0.082	[0.058, 0.107]
(250,100,500)	0.016	[0.014, 0.020]	**(1000,1000,500)**	0.238	[0.199, 0.267]
(250,100,1000)	0.019	[0.011, 0.026]	**(1000,1000,1000)**	0.755	[0.652, 0.829]
(250,200,100)	0.0010	[0.0003, 0.0019]			

Table E.5 MLE Estimation Results for RUM, Three Coverages

	Mixed Logit	
Average β_{1i}	4.89	[4.60, 5.16]
β_2	54.2	[51.6, 56.6]
Mean of ν	0.0017	[0.0016, 0.0017]
SD of ν	0.0007	[0.0007, 0.0007]
Intercept	-2.37	[-2.39, -2.34]
Age	-0.077	[-0.088, -0.066]
Age^2	-0.015	[-0.024, -0.0059]
Female Driver	0.0008	[-0.0098, 0.012]
Single Driver	-0.014	[-0.025, -0.0030]
Married Driver	-0.018	[-0.033, -0.0029]
Credit Score	0.034	[0.023, 0.045]
2+ Drivers	-0.048	[-0.075, -0.020]
Sigma	0.224	[0.209, 0.238]

Table E.6 Average Monetary Loss by Group

	Average Monetary Loss	
All	-50.2	[-52.4, -47.3]
Female Driver	-54.3	[-56.8, -51.0]
Single Driver	-45.1	[-47.1, -42.3]
Young	-45.5	[-47.2, -43.2]
Old	-65.4	[-69.3, -59.4]
Low Credit Driver	-47.6	[-49.3, -44.9]
High Credit Driver	-54.3	[-57.3, -50.3]

References

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, *56*(2), 501-564.

Barseghyan, L., Prince, J., & Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*, *101*(2), 591-631.

Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745-788.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1/2), 122–136.