Supplemental Material
for
“Heterogeneous Choice Sets and Preferences”

Levon Barseghyan
Cornell University

Maura Coughlin
Rice University

Francesca Molinari
Cornell University

Joshua C. Teitelbaum
Georgetown University

February 9, 2021

S1 Theory

S1.1 Unobserved Heterogeneity in Choice Sets as Additively Separable Disturbances

It is possible to represent unobserved heterogeneity in choice sets through additively separable disturbances. In a classic random utility model with $U_i(c) = W_i(c) + \epsilon_{ic}$, one may let $\epsilon_{ic} \in \{-\infty, 0\}$ for each alternative $c \in \mathcal{D}$ and allow ϵ_{ic} to be correlated with $\epsilon_{ic'}$ for any two alternatives $c, c' \in \mathcal{D}$. One would then posit that: if $\kappa = |\mathcal{D}|$ then $\epsilon_{ic} = 0$ for each alternative $c \in \mathcal{D}$; if $\kappa = |\mathcal{D}| - 1$ then $\epsilon_{ic} = -\infty$ for at most one alternative in $\mathcal{D}$ (the identity of which is left unspecified); if $\kappa = |\mathcal{D}| - 2$ then $\epsilon_{ic} = -\infty$ for at most two alternatives in $\mathcal{D}$ (the identities of which are left unspecified); and so forth. This model yields that alternative c is not chosen if $\epsilon_{ic} = -\infty$, which is analogous to alternative c not being chosen when it is not contained in the agent's choice set.

S1.2 Positive Probability of Utility Ties

When utility ties are allowed, one can adapt the definition of $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ as follows:

$$D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \bigcup_{G \subseteq \mathcal{D}: |G| \geqslant \kappa} \left\{ \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \left\{ \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\}, \tag{S1.1}$$

where again the last equality follows from Sen's property α, and now $\arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ may include multiple elements of $\mathcal{D}$ due to the possibility of utility ties. The random closed set $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ contains alternatives up to the $(|\mathcal{D}| - \kappa + 1)$-th best in $\mathcal{D}$, where "best" is defined with respect to $W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$. Due to the possibility of ties, $|D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})|$ may be larger than $|\mathcal{D}| - \kappa + 1$.[1]

To see that our characterization in Theorem 3.1 applied with this new definition of $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ remains sharp, note that the model-implied optimal choice for an agent with attributes $(\mathbf{x}_i, \boldsymbol{\nu}_i)$, utility parameters $\boldsymbol{\delta}$, and choice set G is no longer unique. But this additional multiplicity of optimal choices is incorporated into $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, and all model restrictions continue to be embedded in the requirement that $d_i \in D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, almost surely. The proof of Theorem 3.1 continues to apply, although at the price of additional notation (a selection mechanism that determines the probability with which each optimal choice $d^*_i(G, \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \in \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is selected when multiple alternatives are optimal for a realization G of C_i).

[1] To illustrate, consider the case $|\mathcal{D}| = 5$ and $\kappa = 4$. When utility ties occur with positive probability, for a given $(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ it might be, for example, that three alternatives are tied as first best, and hence at least one of them is in any realization of C_i and $|D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})| = 3$.

S1.3 Computational Simplifications

We omit the subscript i on random variables and random sets throughout this section.

S1.3.1 Sufficient Collection of Test Sets K

Theorem 3.1 and Corollary 3.1 provide a characterization of Θ_I as the collection of $\boldsymbol{\theta} \in \Theta$ that satisfy a finite number of conditional moment inequalities, indexed by the *test sets* $K \subset \mathcal{D}$. In this subsection we provide results to reduce the collection of test sets K for which to check the inequalities from all nonempty proper subsets of $\mathcal{D}$ to a smaller collection.

THEOREM S1.1: *Let the assumptions of Theorem 3.1 hold. Then the following steps yield a sufficient collection of sets K, denoted $\mathbb{K}$, on which to check the inequalities in equation (3.5) to verify if $\boldsymbol{\theta} \in \Theta_I$. Initialize $\mathbb{K} = \{K \subset \mathcal{D} : |K| < \kappa\}$. Then:*

(1) For a given set $K \in \mathbb{K}$, if it holds that $\forall \boldsymbol{\nu} \in \mathcal{V}$ an element of K (possibly different across values of $\boldsymbol{\nu}$) is among the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash K$;[2]

(q) Repeat the following step for $q = 2, \dots, \kappa - 1$. Take any set $K \in \mathbb{K}$ such that $K = K_{q-1} \cup \{c_j\}$ for some K_{q-1} with $|K_{q-1}| = q - 1$ and $\{c_j\} \in \mathbb{K}, K_{q-1} \in \mathbb{K}$ after Steps (1) and (q-1). If $\nexists \boldsymbol{\nu} \in \mathcal{V}$ such that both c_j and at least one element of K_{q-1} are among the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash K$.

If the set D^*_κ does not depend on $\boldsymbol{\delta}$, as in our application in Sections 4–5, the collection $\mathbb{K}$ is invariant across $\boldsymbol{\theta} \in \Theta$.

Proof. Step (1) follows because under the stated condition, $\Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \emptyset) = 1$. Step (q) follows because under the stated condition, the events $\{D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap \{c_j\} \neq \emptyset\}$ and $\{D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K_{q-1} \neq \emptyset\}$ are disjoint. This implies that the right-hand side of the inequality in equation (3.5) is additive, and therefore that inequality evaluated at K is implied by the ones evaluated at $\{c_j\}$ and at K_{q-1}. □

Depending on the structure of the realizations of the random set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$, Theorem S1.1 can be further simplified. The following corollary provides an example.

COROLLARY S1.1: *Let Assumptions 2.1 and 2.2 hold. Suppose all possible realizations of $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ are given by adjacent elements of $\mathcal{D}$, as $\{c_j, c_{j+1}, \dots, c_{j+|\mathcal{D}|-\kappa}\}$, for $j = 1, \dots, \kappa$.*

[2] Here the notation $\mathbb{K} \backslash K$ indicates that the set K is removed from the collection of sets $\mathbb{K}$. In practice, one can implement this step first on sets $K : |K| = 1$, and for K that satisfies the condition remove from $\mathbb{K}$ all sets $K' \supseteq K$. Then repeat the procedure for the remaining sets $K : |K| = 2$, and so forth.

Then the collection of test sets $\mathbb{K}$ *in Theorem S1.1 can be initialized to*

$$\mathbb{K} = \Big\{\{c_1\}, \{c_1, c_2\}, \{c_1, c_2, c_3\}, \cdots, \{c_1, c_2, \ldots, c_{\kappa-1}\},$$
$$\{c_{|\mathcal{D}|}\}, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}\}, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}, c_{|\mathcal{D}|-2}\}, \cdots, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}, \ldots, c_{|\mathcal{D}|-\kappa+2}\}\Big\}, \quad \text{(S1.2)}$$

which contains $2(\kappa - 1)$ *elements.*

Proof. We first establish that if the inequalities in equation (3.5) are satisfied for sets of size $|K| = m$, $m = 1, \ldots, \kappa - 1$, comprised of adjacent alternatives (with respect to $|\mathcal{D}|$), then they are satisfied for all $K \subset \mathcal{D}$.

Let the inequality in equation (3.5) be satisfied for $K_1 = \{c_j, c_{j+1}, \ldots, c_p\}$, for $K_2 = \{c_q, c_{q+1}, \ldots, c_t\}$, with $p < q-1$ so that $K_1 \cap K_2 = \varnothing$, and for $K = K_1 \cup \{c_{p+1}, \ldots, c_{q-1}\} \cup K_2$ (the set that comprises all adjacent alternatives between c_j and c_t). We then show that the inequality for $K_1 \cup K_2$ is redundant. The same argument generalizes to sets comprised of the union of disjoint collections of adjacent alternatives.

Consider first the case that $q - p \geqslant |\mathcal{D}| - \kappa + 1$. Then $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ cannot intersect both K_1 and K_2, and hence

$$P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap (K_1 \cup K_2) \neq \varnothing; \boldsymbol{\gamma}) = P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K_1 \neq \varnothing; \boldsymbol{\gamma}) + P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K_2 \neq \varnothing; \boldsymbol{\gamma})$$

and the result follows.

Consider next the case that $q - p < |\mathcal{D}| - \kappa + 1$. We claim that in this case

$$D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \backslash (K_1 \cup K_2) \neq \varnothing \Rightarrow D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap (K_1 \cup K_2) \neq \varnothing. \quad \text{(S1.3)}$$

To establish this claim, take $c_s \in \{c_{p+1}, \ldots, c_{q-1}\} \equiv K \backslash (K_1 \cup K_2)$. Suppose $c_s \in D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$. Then either $c_p \in D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ or $c_q \in D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$, because $|D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})| = |\mathcal{D}| - \kappa + 1$. The claim follows because $K_1 \cup K_2 \subset K$, and hence $\Pr(d \in K_1 \cup K_2 | \mathbf{x}) \leqslant \Pr(d \in K | \mathbf{x})$, while $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap (K_1 \cup K_2) \neq \varnothing; \boldsymbol{\gamma}) = P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma})$ due to equation (S1.3).

Finally, we show that it suffices to verify equation (3.5) for the sets $K \in \mathbb{K}$ as specified in equation (S1.2). Consider first the case where $|\mathcal{D}| - \kappa + 1 > \kappa - 1$. Then for all $1 < p < q < \kappa$ and $K = \{c_p, c_{p+1}, \ldots, c_q\}$, it holds that $|K| < \kappa - 1$ and, denoting $K^c = \mathcal{D} \backslash K$,

$$\begin{aligned} P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}) &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset K^c; \boldsymbol{\gamma}) \\ &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_1, \ldots, c_{p-1}\}; \boldsymbol{\gamma}) - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_{q+1}, \ldots, c_{\mathcal{D}}\}; \boldsymbol{\gamma}) \\ &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_{q+1}, \ldots, c_{\mathcal{D}}\}; \boldsymbol{\gamma}), \end{aligned} \quad \text{(S1.4)}$$

where the first equality follows by definition, the second follows because $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ is comprised of $|\mathcal{D}| - \kappa + 1$ adjacent alternatives, and the last follows because $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_1, \ldots, c_{p-1}\}; \boldsymbol{\gamma}) = 0$ as $|\{c_1, \ldots, c_{p-1}\}| < \kappa - 1 < |\mathcal{D}| - \kappa + 1$. On the other hand,

$$\Pr(d \in \{c_p, \ldots, c_q\}) \leqslant \Pr(d \in \{c_1, \ldots, c_q\}),$$

and hence if equation (3.5) is satisfied for $K = \{c_1, \ldots, c_q\}$, it is also satisfied for $K = \{c_p, c_{p+1}, \ldots, c_q\}$ for all $1 < p < q < \kappa$. A similar reasoning, with appropriate modifications, holds for sets $K = \{c_{|\mathcal{D}|-q+1}, c_{p+1}, \ldots, c_{|\mathcal{D}|-p+1}\}$.

When $|\mathcal{D}| - \kappa + 1 \leqslant \kappa - 1$, equation (S1.4) continues to hold as stated whenever $p < |\mathcal{D}| - \kappa + 1$. If $p > |\mathcal{D}| - \kappa + 1$, the result follows by the additivity in the second line of equation (S1.4) and the additivity of probabilities, because

$$\Pr(d \in K|\mathbf{x}) \leqslant P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}) \Leftrightarrow \Pr(d \in K^c|\mathbf{x}) \geqslant P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset K^c; \boldsymbol{\gamma}).$$

Hence, the inequality for $K = \{c_p, \ldots, c_q\}$ is implied whenever it is satisfied for $K = \{c_1, \ldots, c_p\}$ and $K = \{c_q, \ldots, c_{|\mathcal{D}|}\}$. □

The following claim establishes that Corollary S1.1 applies when $\boldsymbol{\nu} \in \mathbb{R}$ and the alternatives in the feasible set are vertically differentiated.

CLAIM S1.1: *Let Assumptions 2.1 and 2.2 hold. Let $\mathcal{D} = \{c_1, \ldots, c_{|\mathcal{D}|}\}$ and $\boldsymbol{\nu} = \nu \in \mathbb{R}$. Suppose that: (I) for every pair of alternatives $c_j, c_k \in \mathcal{D}$, $j < k$, and given any $\mathbf{x} \in \mathcal{X}$, there exists a unique threshold $\bar{\nu}_{j,k}(\mathbf{x})$ such that for all $\nu > \bar{\nu}_{j,k}(\mathbf{x})$ alternative c_j has greater utility than alternative c_k and for all $\nu < \bar{\nu}_{j,k}(\mathbf{x})$ alternative c_k has greater utility than alternative c_j; and (II) for every alternative $c_j \in \mathcal{D}$ and given any $\mathbf{x} \in \mathcal{X}$, there exists a $\nu \in \mathbb{R}$ such that c_j is the first best in $\mathcal{D}$. Then, given any $(\mathbf{x}, \nu) \in \mathcal{X} \times \mathbb{R}$ and any $\kappa \geqslant 2$, the set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ comprises adjacent elements of $\mathcal{D}$, as $\{c_j, c_{j+1}, \ldots, c_{j+|\mathcal{D}|-\kappa}\}$, for $j = 1, \ldots, \kappa$.*

Proof. The proof builds on Fact 4 in Barseghyan et al. (2020). Let $|\mathcal{D}| \geqslant 3$ (otherwise the claim holds trivially). Take any $\mathbf{x} \in \mathcal{X}$ and any three alternatives $c_j, c_{j+1}, c_{j+2} \in \mathcal{D}$. Conditions (I) and (II) imply that $\bar{\nu}_{j,j+1}(\mathbf{x}) > \bar{\nu}_{j,j+2}(\mathbf{x}) > \bar{\nu}_{j+1,j+2}(\mathbf{x})$. (In particular, $\bar{\nu}_{j+1,j+2}(\mathbf{x}) > \bar{\nu}_{j,j+2}(\mathbf{x}) > \bar{\nu}_{j,j+1}(\mathbf{x})$ violates condition (II) because c_{j+1} is not first best for any $\nu \in \mathbb{R}$, and every other permutation violates condition (I) due to the transitivity of utility). In other words, the alternatives are *vertically differentiated* in that c_j is first best for all $\nu > \bar{\nu}_{j,j+1}(\mathbf{x})$; c_{j+1} is first best for all $\nu \in (\bar{\nu}_{j+1,j+2}(\mathbf{x}), \bar{\nu}_{j,j+1}(\mathbf{x}))$; and c_{j+2} is first best for all $\nu < \bar{\nu}_{j+1,j+2}(\mathbf{x})$. Consequently, for all $\nu \in \mathbb{R}$, the only possible strict utility rankings of the three alternatives are: $U(c_j) > U(c_{j+1}) > U(c_{j+2})$ (when $\nu > \bar{\nu}_{j,j+1}(\mathbf{x})$);

$U(c_{j+1}) > U(c_j) > U(c_{j+2})$ (when $\bar{\nu}_{j,j+1}(\mathbf{x}) > \nu > \bar{\nu}_{j,j+2}(\mathbf{x})$); $U(c_{j+1}) > U(c_{j+2}) > U(c_j)$ (when $\bar{\nu}_{j,j+2}(\mathbf{x}) > \nu > \bar{\nu}_{j+1,j+2}(\mathbf{x})$); and $U(c_{j+2}) > U(c_{j+1}) > U(c_j)$ (when $\nu < \bar{\nu}_{j+1,j+2}(\mathbf{x})$). Thus, alternative c_{j+1} is never the third best among the three alternatives. This implies that if c_j and c_{j+2} both have greater utility than a fourth alternative c_m, $m \notin \{j, j+1, j+2\}$, then c_{j+1} also has greater utility than c_m. It follows that for any $(\mathbf{x}, \nu) \in \mathcal{X} \times \mathbb{R}$, the set $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ comprises adjacent elements of $\mathcal{D}$, as $\{c_j, c_{j+1}, \ldots, c_{j+|\mathcal{D}|-\kappa}\}$, for $j = 1, \ldots, \kappa$. □

When Assumption 3.1 is maintained, the logic of Theorem S1.1 can be used to obtain a collection of sufficient test sets K on which to verify the inequalities in (3.7), by applying its Steps 2.1-2.$(\kappa - 1)$ to the random sets $D^*_q(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$, $q = \kappa, \ldots, |\mathcal{D}|$. Further simplifications are possible when interest centers on specific projections of Θ_I, using the fact that $D^*_{q+1}(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \subset D^*_q(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ for all $q \geqslant \kappa$. As discussed following Corollary 3.1, when Assumption 3.1 is maintained the projection of Θ_I on $[\boldsymbol{\delta}; \boldsymbol{\gamma}]$ is obtained by setting $\pi_\kappa(\mathbf{x}; \boldsymbol{\eta}) = 1$ and $\pi_q(\mathbf{x}; \boldsymbol{\eta}) = 0$, $q = \kappa + 1, \ldots, |\mathcal{D}|$. Hence, Steps 2.1-2.$(\kappa - 1)$ in Theorem S1.1 applied only to $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ deliver the sufficient collection of sets K on which to verify (3.7) to obtain the sharp identification region for $[\boldsymbol{\delta}; \boldsymbol{\gamma}]$. On the other hand, the projection of Θ_I on $\pi_q(\mathbf{x}; \boldsymbol{\eta})$, $q = \kappa + 1, \ldots, |\mathcal{D}|$, is obtained by setting $\pi_l(\mathbf{x}; \boldsymbol{\eta}) = 0$ for all $l \notin \{q, \kappa\}$, and that on $\pi_\kappa(\mathbf{x}; \boldsymbol{\eta})$ by setting $\pi_l(\mathbf{x}; \boldsymbol{\eta}) = 0$ for all $l = \kappa + 2, \ldots, |\mathcal{D}|$. Hence, Steps 2.1-2.$(\kappa - 1)$ in Theorem S1.1 applied, respectively, to only $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ and $D^*_q(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ deliver the sufficient collection of sets K on which to verify (3.7) to obtain the sharp identification region for π_q, $q = \kappa + 1, \ldots, |\mathcal{D}|$, and applied only to $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ and $D^*_{\kappa+1}(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ deliver the sufficient collection of sets K on which to verify (3.7) to obtain the sharp identification region for π_κ.

The two corollaries that follow illustrate the specific adaptations of Theorem S1.1 that we use in our application in Sections 4–5. Proofs are omitted because the corollaries follow immediately from Theorem S1.1.

Corollary S1.2: *Let $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 3$. Suppose that all assumptions in Corollary 3.1 hold and that $\boldsymbol{\nu} = \nu \in \mathbb{R}$ with support $[0, \bar{\nu}]$, $\bar{\nu} < \infty$. Then the following steps yield a sufficient collection of sets K, denoted $\mathbb{K}$, on which to check the inequalities in equation (3.7) to obtain sharp bounds on π_5. Initialize $\mathbb{K} = \{K : K \subset \mathcal{D}\}$. Then:*

1. *For any set $K = \{c_j, c_k\} \subset \mathcal{D}$, if $\nexists \nu \in [0, \bar{\nu}]$ such that both c_j and c_k are among the best 3 alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash \{c_j, c_k\}$;*

2. *Set $\mathbb{K} = \mathbb{K} \backslash \{c_j, c_k, c_l\}$ for all $j, k, l \in \{1, 2, 3, 4, 5\}$.*

Corollary S1.3: *Let $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 3$. Suppose that all assumptions in Corollary 3.1 hold and that $\boldsymbol{\nu} = \nu \in \mathbb{R}$ with support $[0, \bar{\nu}]$, $\bar{\nu} < \infty$. Then the following* *steps yield a sufficient collection of sets K, denoted $\mathbb{K}$, on which to check the inequalities in equation (3.7) to obtain sharp bounds on π_4. Initialize $\mathbb{K} = \{K : K \subset \mathcal{D}\}$. Then:*

1. *For any set $K = \{c_j, c_k\} \subset \mathcal{D}$, if $\nexists \nu \in [0, \bar{\nu}]$ such that both c_j and c_k are among the best 3 alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash \{\{c_j, c_k\}, \{\mathcal{D} \backslash \{c_j, c_k\}\}\}$;*

2. *For any set $K = \{c_j, c_k, c_l\} \subset \mathcal{D}$ such that $\{c_j, c_k\} \in \mathbb{K}$ after Step 1, if $\nexists \nu \in [0, \bar{\nu}]$ such that both c_l and at least one element of $\{c_j, c_k\}$ are among the best 3 alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash \{c_j, c_k, c_l\}$;*

3. *For any set $K \in \mathbb{K}$, if $\forall \nu \in [0, \bar{\nu}]$ one element of K, possibly different across values of ν, is among the best 2 alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash K$.*

In our application in Sections 4–5, the number of inequalities obtained through application of the foregoing results (taking into account the 65 hypercubes on $(\mu, \bar{p})$) is $6 \times 65 = 390$ for the sharp identification region of $\boldsymbol{\gamma}$; $17 \times 65 = 1{,}105$ for the sharp identification region of π_5; and $15 \times 65 = 975$ for the sharp identification region of π_4.

S1.4 An Equivalent Characterization Based on Convex Optimization

The characterization in Theorem 3.1 can equivalently be written in terms of a convex optimization problem.

Corollary S1.4: *Let Assumptions 2.1 and 2.2 hold and let $\Theta = \Delta \times \Gamma$. Then*

$$\Theta_I = \left\{ \theta \in \Theta : \max_{\mathbf{u} \in \mathbb{R}^{|\mathcal{D}|} : ||\mathbf{u}|| \leqslant 1} \left[\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \int_{\boldsymbol{\tau} \in \mathcal{V}} \max_{d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta})} \left(\mathbf{u}^\top \mathbf{q}^{d^*} \right) dP(\boldsymbol{\tau}; \boldsymbol{\gamma}) \right] = 0, \mathbf{x} - a.s. \right\}, \tag{S1.5}$$

where $\mathbf{p}(\mathbf{x}) = [\Pr(d = c_1 | \mathbf{x}) \ \ldots \ \Pr(d = c_{|\mathcal{D}|} | \mathbf{x})]^\top$ *and, for a given* $d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$, $\mathbf{q}^{d^*} = [\mathbf{1}(d^* = c_1) \ \ldots \ \mathbf{1}(d^* = c_{|\mathcal{D}|})]^\top$.

Proof. We establish the equivalence between equations (3.5) in the paper and (S1.5) here.[3] Due to the positive homogeneity in $\mathbf{u}$ of $\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \int_{\boldsymbol{\tau} \in \mathcal{V}} \max_{d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta})} \mathbf{u}^\top \mathbf{q}^{d^*} dP(\boldsymbol{\tau}; \boldsymbol{\gamma})$, we have that

$$\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \int_{\boldsymbol{\tau} \in \mathcal{V}} \max_{d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta})} \mathbf{u}^\top \mathbf{q}^{d^*} dP(\boldsymbol{\tau}; \boldsymbol{\gamma}) \leqslant 0 \tag{S1.6}$$

holds for all $\mathbf{u} : ||\mathbf{u}|| \leqslant 1$ if and only if expression (S1.6) holds for all $\mathbf{u} \in \mathbb{R}^{|\mathcal{D}|}$. Consider the specific subset of vectors $\mathsf{U} = \{\mathbf{u} \in \mathbb{R}^{|\mathcal{D}|} : u_j \in \{0, 1\}, j = 1, \ldots, |\mathcal{D}|\}$. Each vector $\mathbf{u} \in \mathsf{U}$

[3] The argument of proof goes through similar steps as in Molchanov and Molinari (2018, Theorem 3.28).

uniquely corresponds to a subset $K_{\mathbf{u}} = \{c_1 u_1, \ldots, c_{|\mathcal{D}|} u_{|\mathcal{D}|}\}$. For a given $\mathbf{u}$, $\mathbf{u}^\top \mathbf{q}^{d^*} = 1$ if $d^* \in K_{\mathbf{u}}$ and $\mathbf{u}^\top \mathbf{q}^{d^*} = 0$ otherwise. Hence, the corresponding inequality in (S1.6) reduces to

$$\Pr(d \in K_{\mathbf{u}}|\mathbf{x}) = \mathbf{u}^\top \mathbf{p}(\mathbf{x}) \leqslant \mathrm{E}\left[\max_{d^* \in D^*_\kappa(\mathbf{x},\boldsymbol{\tau};\boldsymbol{\delta})} \mathbf{u}^\top \mathbf{q}^{d^*} |\mathbf{x};\boldsymbol{\gamma}\right] = P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K_{\mathbf{u}} \neq \varnothing;\boldsymbol{\gamma}).$$

It then follows that any θ in the set defined in equation (S1.5) belongs to the set defined in equation (3.5) because $\{K : K \subseteq \mathcal{D}\} = \{K_{\mathbf{u}} : \mathbf{u} \in \mathsf{U}\}$.

Conversely, take a θ in the set defined by equation (3.5). Then, by Theorem A.1, there exists a selection d^* of $D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})$ such that for all $c \in \mathcal{D}$ and $\mathbf{x} - a.s.$, $\Pr(d = c|\mathbf{x}_i) = \Pr(d^* = c|\mathbf{x}_i)$. Hence, θ belongs to the set defined in equation (S1.5). □

As the set $D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})$ is comprised of the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, it can have only a finite number of realizations, as discussed in Section 3.4, which we denote $D^1, \ldots, D^h$. Hence, the characterization in equation (S1.5) can be rewritten as

$$\Theta_I = \left\{\theta \in \Theta : \max_{\mathbf{u} \in \mathbb{R}^{|\mathcal{D}|}:||\mathbf{u}|| \leqslant 1} \left[\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \sum_{j=1}^{h} \left(\max_{d^* \in D^j} \mathbf{u}^\top \mathbf{q}^{d^*}\right) P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) = D^j;\gamma)\right] = 0, \mathbf{x} - a.s.\right\}.$$

This means that to determine whether a given $\theta \in \Theta$ belongs to Θ_I, it suffices to maximize an easy-to-compute superlinear, hence concave, function over a convex set, and check if the resulting objective value vanishes. Several efficient algorithms in convex programming are available to solve this problem; see, for example, the Matlab software for disciplined convex programming CVX (Grant and Boyd 2010).

S1.5 Additively Separable Extreme Value Type 1 Unobserved Heterogeneity

We now explain how to compute $P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K \neq \varnothing;\boldsymbol{\gamma})$ when $\boldsymbol{\nu} = (\boldsymbol{\upsilon}, (\epsilon_c, c \in \mathcal{D}))$ and $W(\mathbf{x}_c,\boldsymbol{\nu};\boldsymbol{\delta}) = \omega(\mathbf{x}_c,\boldsymbol{\upsilon};\boldsymbol{\delta}) + \epsilon_c$, with ϵ_c independently and identically distributed Extreme Value Type 1 *and* independent of $\boldsymbol{\upsilon}$, as in a mixed logit (McFadden and Train 2000).

Given a realization G of the choice set and $\tilde{c} \in G$ (and no utility ties), we have

$$\begin{aligned}\Pr(d^*(G,\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) = \tilde{c}|\mathbf{x},\boldsymbol{\upsilon}) &= \Pr(W(\mathbf{x}_{\tilde{c}},\boldsymbol{\nu};\boldsymbol{\delta}) \geqslant W(\mathbf{x}_c,\boldsymbol{\nu};\boldsymbol{\delta})\ \forall c \in G|\boldsymbol{\upsilon}) \\ &= \frac{\exp(\omega(\mathbf{x}_{\tilde{c}},\boldsymbol{\upsilon};\boldsymbol{\delta}))}{\sum_{c \in G} \exp(\omega(\mathbf{x}_c,\boldsymbol{\upsilon};\boldsymbol{\delta}))}. \end{aligned} \tag{S1.7}$$

Conditional on $\boldsymbol{\upsilon}$, one can leverage the closed-form expressions in equation (S1.7) to compute $P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K \neq \varnothing;\boldsymbol{\gamma})$ so that numerical integration is needed *only* for $\boldsymbol{\upsilon}$. The same result applies, with q replacing κ, to compute $P(D^*_q(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K \neq \varnothing;\boldsymbol{\gamma})$ in Corollary 3.1.

Theorem S1.2: *Suppose that* $\boldsymbol{\nu} = (\boldsymbol{\upsilon}, (\epsilon_c, c \in \mathcal{D}))$ *and* $W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}) = \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta}) + \epsilon_c$, *with* ϵ_c *independently and identically distributed Extreme Value Type 1* and *independent of* $\boldsymbol{\upsilon}$*. Conditional on* $\boldsymbol{\upsilon}$*, any* $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ *can be computed as a linear combination over different sets* G *of expression* (S1.7)*. Hence, any* $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma})$ *can be computed as an integral with respect to the distribution of* $\boldsymbol{\upsilon}$ *of linear combinations over different sets* G *of expression* (S1.7)*.*

To prove this theorem, we first establish two auxiliary results. The first one states that the probability of at least one alternative in K being preferred to all alternatives in $\mathcal{D} \backslash K$ is the sum over all elements of K that each is first best in $\mathcal{D}$.

Claim S1.2: *Conditional on* $\boldsymbol{\upsilon}$*, the probability that at least one alternative in a set* $K \subset \mathcal{D}$ *is better than all alternatives in the set* $\mathcal{D} \backslash K$ *is given by*

$$\Pr(\vee_{c' \in K}\ W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})\ \forall c \in \mathcal{D} \backslash K | \boldsymbol{\upsilon}) = \sum_{c' \in K} \frac{\exp(\omega(\mathbf{x}_{c'}, \boldsymbol{\upsilon}; \boldsymbol{\delta}))}{\sum_{c \in \mathcal{D}} \exp(\omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta}))}.$$

Proof of Claim S1.2. We first establish equivalence of the following events:

$$\{\exists c' \in K\ s.t.\ W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta});\ \forall c \in \mathcal{D} \backslash K\} \\ \iff \cup_{c' \in K} \{W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}), \forall c \in \mathcal{D} \backslash c'\}. \quad \text{(S1.8)}$$

The right-to-left implication in (S1.8) is immediate. The left-to-right implication can be established by contradiction, observing that the complement of the event in the right-hand side of (S1.8) is that there exists a $c \in \mathcal{D} \backslash K$ that is preferred to all other alternatives. The result then follows because the events in the right-hand side of (S1.8) are disjoint. □

Next, recall that, as discussed in Section 3.4, the set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ can only take on a finite number of realizations, denoted $D^1, \dots, D^h$, with $|D^j| = |\mathcal{D}| - \kappa + 1$ for all $j = 1, \dots, h$. We show how to compute the probability of any of these realizations.

Claim S1.3: *For each* $j = 1, \dots, h$, $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) = D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ *can be computed as a linear combination of expression* (S1.7) *for different sets* G*.*

Proof of Claim S1.3. Note that

$$P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) = D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma}) = P(W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}),\ \forall c' \in D^j, \forall c \in \mathcal{D} \backslash D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma}).$$

Given this, the proof proceeds sequentially. Suppose $|D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})| = 1$. Then the result follows immediately (with $G = \mathcal{D}$). Suppose $|D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})| = 2$. Then we have $D^j = \{c', c''\}$

for some $c', c'' \in \mathcal{D}$, and

$$\begin{aligned}
&P(\{W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})\} \cap \{W(\mathbf{x}_{c''}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})\} \; \forall c \in \mathcal{D} \backslash D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma}) \\
&= P(W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}) \; \forall c \in \mathcal{D} \backslash D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma}) + P(W(\mathbf{x}_{c''}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta} | \boldsymbol{\upsilon}; \boldsymbol{\gamma}) \; \forall c \in \mathcal{D} \backslash D^j) \\
&\quad - P(\{W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})\} \cup \{W(\mathbf{x}_{c''}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})\} \; \forall c \in \mathcal{D} \backslash D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma}).
\end{aligned}$$

The first term in this expression can be computed by applying equation (S1.7) with $G = \mathcal{D} \backslash c''$; the second term can be computed by applying equation (S1.7) with $G = \mathcal{D} \backslash c'$; the last term, by Claim S1.2, can be computed as the sum over $\tilde{c} \in D^j$ of equation (S1.7) with $G = \mathcal{D}$.

For $|D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})| \geqslant 3$ one can proceed iteratively using the inclusion/exclusion formula and applying Claim S1.2. □

With these results in hand, we prove Theorem S1.2.

Proof of Theorem S1.2. By Claim S1.3 we can compute $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) = D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ for each D^j such that $|D^j| = |\mathcal{D}| - \kappa + 1$ as a linear combination of expression (S1.7) with different sets G. To obtain the result in Theorem S1.2, for each set K one can simply sum $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) = D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ over the sets D^j such that $D^j \cap K \neq \emptyset$. □

S2 Additional Details on Statistical Inference

As explained in Section 5, we base our confidence sets for the vector $\boldsymbol{\theta}$ on the Kolmogorov-Smirnov test statistic suggested by Andrews and Shi (2013, equation (3.7) on p. 618) [hereafter, AS], which in our framework simplifies to

$$T_n(\boldsymbol{\theta}) = n \max_{j=1,\ldots,J; K \in \mathbb{K}} \max \left\{ \frac{\bar{m}_{n,K,j}(\boldsymbol{\theta})}{\hat{\sigma}_{n,K,j}(\boldsymbol{\theta})}, 0 \right\}^2$$

where $\bar{m}_{n,K,j}(\boldsymbol{\theta})$ and $\hat{\sigma}_{n,K,j}(\boldsymbol{\theta})$ are defined in Section 5. Our application of the method proposed by AS computes bootstrap-based critical values to obtain a confidence set

$$CS = \{\boldsymbol{\theta} \in \Theta : T_n(\boldsymbol{\theta}) \leqslant \hat{c}_{n,1-\alpha+\xi}(\boldsymbol{\theta}) + \xi\}$$

where $\xi > 0$ is an arbitrarily small constant which we set equal to 10^{-6} as suggested by AS (p. 625). In practice, we evaluate $T_n(\boldsymbol{\theta})$ and the bootstrap-based critical value $\hat{c}_{n,1-\alpha+\xi}(\boldsymbol{\theta})$ on a grid of values of $\boldsymbol{\theta}$ designed to give good coverage of the $(\mathrm{E}(\nu), \mathrm{Var}(\nu))$-space to obtain a precise description of the confidence set for this pair of parameters. To explain how this grid is constructed, we note that given the assumption that $\nu_i \sim Beta(\gamma_1, \gamma_2)$ with support $[0, 0.03]$,

$\mathrm{E}(\nu) \in 0.03 \times (0,1]$ and $\mathrm{Var}(\nu) \in 0.0009 \times (0, 0.25]$. We therefore obtain a grid of values over (γ_1, γ_2) comprised of 665,603 points, such that the associated grid on $(\mathrm{E}(\nu), \mathrm{Var}(\nu))$ has first coordinate in $0.03 \times [0.0005, 0.9995]$ with step size 0.03×0.0005, and second coordinate in $0.0009 \times (0.0005, 0.25]$ with step size 0.0009×0.0005.[4] The approximation of $\hat{c}_{n,1-\alpha+\xi}(\boldsymbol{\theta})$ is based on the bootstrap procedure detailed in AS (Section 9) and uses 1,000 bootstrap replications.[5] The procedure takes as inputs a *GMS function* φ, a *GMS sequence* τ_n such that $\tau_n \to \infty$ as $n \to \infty$, and a *non-decreasing sequence of positive constants* β_n such that $\beta_n / \tau_n \to 0$ as $n \to \infty$, which together are used to determine which moment inequalities are sufficiently close to binding to contribute to the limiting distribution of $T_n(\boldsymbol{\theta})$. We use the GMS function proposed by AS (equation (4.10) on p. 627):[6]

$$\varphi_{K,j}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \tau_n^{-1} \sqrt{n} \bar{m}_{n,K,j}(\boldsymbol{\theta}) / \hat{\sigma}_{n,K,j}(\boldsymbol{\theta}) \geqslant -1 \\ -\beta_n & \text{otherwise,} \end{cases}$$

and we set $\tau_n = (0.3 \ln n)^{1/2}$ and $\beta_n = (0.4 \ln n / \ln \ln n)^{1/2}$ as recommended by AS (p. 643).

Similar to AS, the KMS procedure takes as inputs a GMS function φ and a GMS sequence τ_n.[7] To simplify computations, we use the *hard threshold* GMS function:[8]

$$\varphi_{K,j}(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \tau_n^{-1} \sqrt{n} \bar{m}_{n,K,j}(\boldsymbol{\theta}) / \hat{\sigma}_{n,K,j}(\boldsymbol{\theta}) \geqslant -1 \\ -\infty & \text{otherwise.} \end{cases}$$

The procedure also requires a regularization parameter $\rho \geqslant 0$, which (like φ and τ_n) enters the calibration of $\hat{c}^f_{n,1-\alpha}$ and introduces a conservative distortion that is required to obtain uniform coverage of projections. The smaller is the value of ρ, the larger is the conservative distortion, but the higher is the confidence that the critical value is uniformly valid in situations where the local geometry of Θ_I makes inference especially challenging. For a discussion, see KMS (Section 2.2). We choose the value of ρ as follows. We begin with the recommendation in KMS (Section 2.4). To further guard against possible irregularities in the local geometry of Θ_I, we reduce the resulting value of ρ by 20 percent.

[4] To obtain confidence intervals on π_5, π_4, and π_3, we first evaluate $T_n(\boldsymbol{\theta})$ on a coarser grid and compare it with the AS critical value. For each π_q, $q = 3, 4, 5$, we then refine the grid around the extreme values of π_q that are not rejected, for a final step size of 0.01 on π_q and 0.05 on each component of (γ_1, γ_2).

[5] Compared to the description in AS (Section 9), note that our moment inequalities are of the $\leqslant$ form, whereas AS's are of the $\geqslant$ form.

[6] AS label the GMS sequence κ_n, but we use τ_n to avoid confusion with our use of κ for the (known and fixed) minimum choice set size in Assumption 2.2.

[7] Our findings based on the AS and KMS methods are robust to the choice of tuning parameters, as indicated by results available from the authors upon request.

[8] This function was proposed by Andrews and Soares (2010) and labeled $\varphi^{(1)}$ on p. 131 of their article.

S3 Additional Results

S3.1 Claim Probabilities

The claim probabilities originate from Barseghyan et al. (2018). We estimate the households' claim probabilities using the company's claims data. We assume that household i's auto collision claims in year t follow a Poisson distribution with mean λ_{it}. We also assume that the household's deductible choice does not influence its claim rates λ_{it} (Assumption 4.1(II)). We treat the household's claim rate as a latent random variable and assume that $\ln \lambda_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + \varepsilon_i$, where $\mathbf{X}_{it}$ is a vector of observables and $\exp(\varepsilon_i)$ follows a Gamma distribution with unit mean and variance ϕ. We perform a Poisson panel regression with random effects to obtain maximum likelihood estimates of $\boldsymbol{\beta}$ and ϕ. In an effort to obtain the most precise estimates, we use the full set of auto collision claims data, which comprises 1,349,853 household-year records. For each household, we calculate a fitted claim rate $\widehat{\lambda}_i$ conditional on the household's observables at the time of first purchase and its subsequent claims experience. More specifically, $\widehat{\lambda}_i = \exp(\mathbf{X}_i'\widehat{\boldsymbol{\beta}})\,\mathrm{E}(\exp(\varepsilon_i)|\mathbf{Y}_i)$, where $\mathbf{Y}_i$ records household i's claims experience after purchasing the policy and $\mathrm{E}(\exp(\varepsilon_i)|\mathbf{Y}_i)$ is calculated using the maximum likelihood estimate of ϕ. In principle, a household may experience one or more claims during the policy period. We assume that households disregard the possibility of experiencing more than one claim (Assumption 4.1(I)). Given this, we transform $\widehat{\lambda}_i$ into a claim probability $\mu_i \equiv 1 - \exp(-\widehat{\lambda}_i)$, which follows from the Poisson probability mass function, and round it to the nearest half percentage point. We treat μ_i as data.

S3.2 Deductible Choices

Table S3.1 reports the sample distribution of deductible choices by octiles of base price $\bar{p}_i$ and claim probability μ_i. The octiles are the hypercubes referenced in Sections 5 and S2 (other than the one that contains all households).

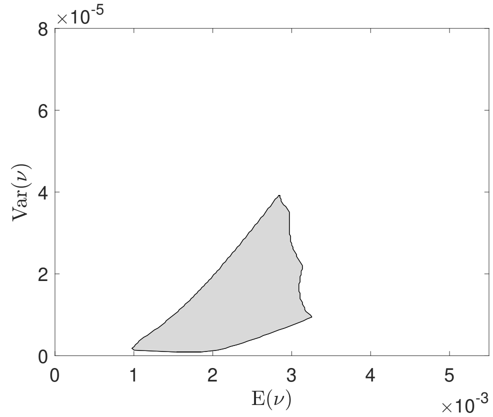
S3.3 Subgroup Results

Figure S3.1 depicts the AS 95 percent confidence set for $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ for population subgroups based on gender, age, and insurance score of the principal driver. In addition, Table S3.2 reports (i) the KMS 95 percent confidence interval for the mean of ν_i and (ii) 95 percent confidence intervals for the 25th and 75th percentiles of ν_i based on projections of the AS confidence set. For the mean, we report the actual confidence interval as well as the risk premium, for a lottery that yields a loss of \$1000 with probability 10 percent, implied

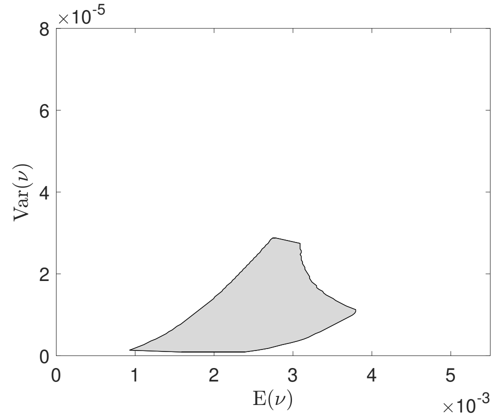
Table S3.1: Deductible Choices by Octiles of $\bar{p}$ and μ

$\bar{p}$ octile	μ octile	Obs.	Percent choosing deductible $100	$200	$250	$500	$1000
1	1	2,756	3.3	31.2	18.9	43.8	2.9
1	2	2,901	3.6	31.8	18.7	43.6	2.2
1	3	2,661	2.9	32.1	20.0	43.6	1.5
1	4	2,113	3.4	34.2	20.6	40.8	1.0
1	5	2,116	3.9	32.1	20.2	42.2	1.5
1	6	1,630	4.2	34.5	21.9	38.9	0.6
1	7	1,233	4.4	34.1	22.8	38.7	0.0
1	8	660	5.0	39.4	25.6	30.0	0.0
2	1	1,949	1.0	20.8	17.0	57.1	4.0
2	2	1,944	2.0	22.3	16.9	56.4	2.5
2	3	1,543	1.9	25.7	19.1	50.7	2.6
2	4	2,152	2.0	23.1	18.5	54.4	2.0
2	5	1,320	2.3	26.7	18.0	50.8	2.2
2	6	1,979	1.6	25.6	20.1	51.1	1.6
2	7	1,584	1.8	26.5	22.6	47.9	1.3
2	8	1,151	2.0	26.5	22.7	48.7	0.2
3	1	1,362	0.7	20.4	14.3	59.8	4.7
3	2	1,914	0.8	18.5	14.6	62.1	3.9
3	3	2,127	0.8	19.8	16.1	60.0	3.2
3	4	1,518	1.3	20.3	17.7	59.4	1.4
3	5	2,255	1.0	19.9	17.6	59.4	2.1
3	6	1,773	0.8	19.9	18.4	59.1	1.9
3	7	1,729	1.2	21.1	20.0	56.7	1.1
3	8	1,602	1.2	20.7	22.2	54.9	0.9
4	1	1,340	0.7	12.7	13.7	67.5	5.3
4	2	1,458	0.8	14.1	15.2	65.8	4.3
4	3	1,632	0.7	15.1	15.4	66.1	2.8
4	4	1,595	0.6	14.7	16.6	64.8	3.3
4	5	1,606	0.8	14.3	17.1	65.4	2.5
4	6	1,705	0.6	16.1	15.2	65.5	2.6
4	7	1,974	0.7	15.4	17.0	65.5	1.5
4	8	1,914	1.0	17.3	17.7	62.8	1.2
5	1	1,126	0.4	11.4	12.6	70.5	5.2
5	2	1,547	0.1	11.8	11.9	71.7	4.5
5	3	1,609	0.5	10.4	13.0	71.6	4.5
5	4	1,522	0.5	10.6	14.5	71.4	3.0
5	5	2,066	0.7	10.8	12.8	72.1	3.5
5	6	1,697	0.6	12.5	14.7	69.2	2.9
5	7	1,801	0.2	12.2	14.6	70.9	2.2
5	8	2,128	0.5	11.9	17.1	68.8	1.6
6	1	1,303	0.3	6.7	9.1	78.3	5.6
6	2	1,403	0.2	6.9	11.4	75.5	6.0
6	3	1,326	0.5	7.3	11.2	76.8	4.2
6	4	1,784	0.3	8.1	11.2	76.2	4.2
6	5	1,589	0.2	7.9	9.8	78.0	4.1
6	6	1,725	0.5	8.9	12.0	74.7	3.9
6	7	2,061	0.1	7.3	11.2	78.4	3.1
6	8	2,363	0.1	9.0	12.3	76.3	2.2
7	1	1,521	0.3	5.2	6.9	81.1	6.5
7	2	1,351	0.1	5.6	7.5	80.1	6.7
7	3	1,665	0.2	4.1	8.6	80.2	6.8
7	4	1,646	0.1	5.0	6.7	81.7	6.4
7	5	1,726	0.1	5.0	7.4	82.6	5.0
7	6	1,865	0.1	4.9	7.9	82.5	4.6
7	7	2,045	0.1	5.7	7.6	82.4	4.2
7	8	2,452	0.2	5.4	9.1	81.0	4.4
8	1	2,636	0.0	1.3	2.5	74.2	21.9
8	2	1,553	0.1	1.5	1.8	80.3	16.4
8	3	1,463	0.0	1.6	3.1	82.8	12.4
8	4	1,568	0.0	1.4	2.7	80.2	15.6
8	5	1,384	0.0	1.8	2.0	80.6	15.6
8	6	1,570	0.1	2.0	3.0	78.9	16.1
8	7	1,501	0.0	1.2	2.5	82.7	13.7
8	8	1,698	0.1	2.1	3.3	81.0	13.5

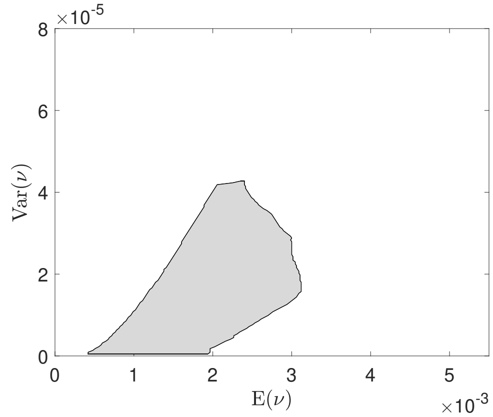
Notes: Analysis sample (111,890 households).



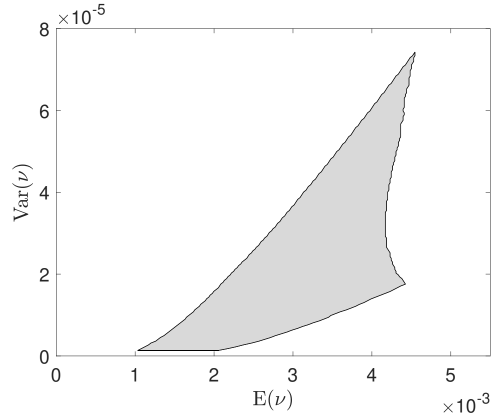
(a) Male



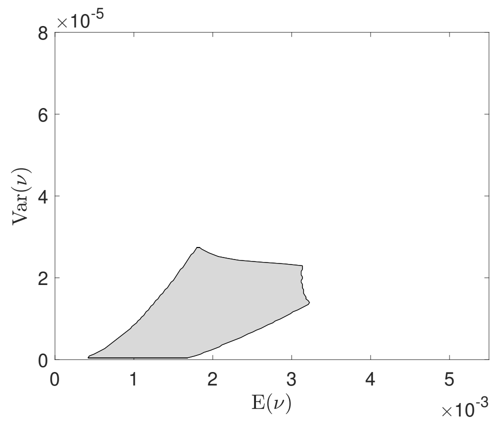
(b) Female



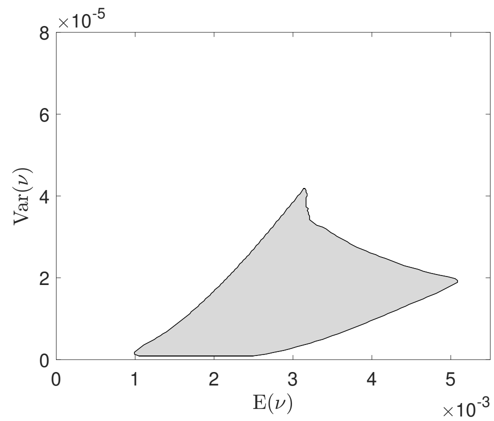
(c) Young



(d) Old



(e) Low insurance score



(f) High insurance score

Figure S3.1: AS 95 percent confidence sets for $(\mathrm{E}(\nu), \mathrm{Var}(\nu))$.

Table S3.2: Distribution of Absolute Risk Aversion

	Mean		Implied risk premium					
			Mean		25th pctl.		75th pctl.	
	LB	UB	LB	UB	LB	UB	LB	UB
Male	0.00104	0.00321	$ 61	$279	$ 0	$ 73	$ 76	$426
Female	0.00101	0.00377	$ 59	$339	$ 0	$117	$ 81	$485
Young	0.00044	0.00306	$ 22	$263	$ 0	$ 95	$ 0	$407
Old	0.00107	0.00432	$ 63	$393	$ 0	$ 73	$ 95	$548
Low insurance score	0.00042	0.00315	$ 21	$273	$ 0	$ 73	$ 7	$425
High insurance score	0.00102	0.00501	$ 60	$452	$ 0	$127	$ 85	$591

Notes: 95 percent confidence intervals. LB = lower bound. UB = upper bound. Implied risk premia for a lottery that yields a loss of $1000 with probability 10 percent.

by each bound. For the percentiles, we report only the implied risk premia. For the most part, the subgroup results are comparable to the results for all households. The notable exceptions are the lower bounds on the mean for households with young principal drivers and households with low insurance scores. These lower bounds are on the order of $4 \cdot 10^{-4}$ (which implies a risk premium of about $20), whereas the corresponding lower bounds for the other subgroups and the population are on the order of 10^{-3} (which implies a risk premium of about $60).[9] Strikingly, the lower bounds on the 75th percentile for these two subgroups correspond to risk premia of 17 cents and $7, respectively.

Table S3.3 reports KMS 95 percent confidence intervals for π_5, π_4, and π_3 for the same population subgroups. The interesting quantities are the upper bounds on π_5 and π_4. The former is the maximum fraction of households whose deductible choices can be rationalized with full size choice sets, while the latter is the maximum fraction of households whose deductible choices can be rationalized with full-1 choice sets.[10] We find, inter alia, that: (i) at least 70 percent of households with female principal drivers require limited choice sets to explain their deductible choices, whereas at least 74 percent of households with male principal drivers require limited choice sets; (ii) at least 73 percent of households with old principal drivers require limited choice sets to explain their deductible choices, whereas at least 75 percent of households with young principal drivers require limited choice sets; and (iii) at least 67 percent of households with low insurance scores require limited choice sets to explain their deductible choices, whereas at least 73 percent of households with high insurance scores require limited choice sets.[11]

[9] Because the subgroups all have different confidence sets (as well as different sample sizes), it is possible that a result for all households is not a weighted average of the corresponding results within a subgroup.

[10] With $\kappa = 3$, the lower bounds on π_5 and π_4 are zero, the lower bound on π_3 is one minus the upper

Table S3.3: Distribution of Choice Set Size

	π_5 (full)		π_4 (full-1)		π_3 (full-2)	
	LB	UB	LB	UB	LB	UB
Male	0.00	0.26	0.00	0.85	0.15	1.00
Female	0.00	0.30	0.00	0.90	0.10	1.00
Young	0.00	0.25	0.00	1.00	0.00	1.00
Old	0.00	0.27	0.00	0.96	0.04	1.00
Low insurance score	0.00	0.33	0.00	1.00	0.00	1.00
High insurance score	0.00	0.27	0.00	1.00	0.00	1.00

Notes: KMS 95 percent confidence intervals. LB = lower bound. UB = upper bound.

S3.4 Admissible Probability Density Functions

Figure S3.2 depicts a 95 percent confidence set for an outer region of admissible probability density functions of ν_i for all households. To construct the outer region (shaded in grey), we find at each point on a grid of 101 values of ν_i the minimum and maximum values of all probability density functions implied by values of $\boldsymbol{\theta}$ in the AS 95 percent confidence set. This gives us 101 points on the lower and upper envelopes of admissible probability density functions. In other words, we obtain pointwise sharp lower and upper bounds on the set of admissible probability density functions. Although the bounds are pointwise sharp, the region is labeled an outer region because not all probability density functions in it are consistent with the distribution of observed choices. The figure also superimposes the predicted density functions of ν_i based on point estimates obtained under the UR and ASR models. The UR predicted density function does not lie entirely inside the confidence set, whereas the AR predicted density function does (although we note that this does not necessarily imply that the true choice formation process is an ASR process).

S3.5 Suboptimal Choices

As we state in Section 5.2.1, with full size choice sets, our model cannot explain the frequency of the \$200 deductible in our data. The reason is that, with full size choice sets, our model satisfies the following conditional rank order property, which is a generalization of the rank order property established by Manski (1975) for random utility models that are linear in the nonrandom parameters and feature an additive i.i.d. disturbance in the utility function.

bound on π_4, and the upper bound on π_3 is one.

[11] Because the subgroups all have different confidence sets (as well as different sample sizes), it is possible that the upper bound on π_5 for all households is not a weighted average of the upper bounds on π_5 within a subgroup. The same is true for the upper bound on π_4 (and, therefore, for the lower bound on π_3).

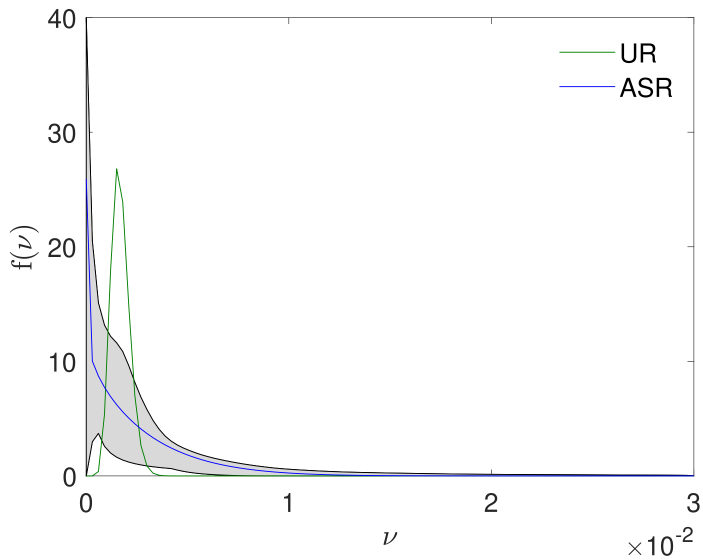


Figure S3.2: Confidence set for outer region of admissible probability density functions of ν.

Notes: The figure depicts a 95 percent confidence set for an outer region of admissible probability density functions of ν_i. It also superimposes the implied probability density functions of ν_i based on point estimates obtained under the UR and ASR models.

PROPERTY S3.1 (Conditional Rank Order Property): *For all* $c, c' \in \mathcal{D}$*,* $\Pr(d = c'|\mathbf{x}, \boldsymbol{\nu}) \geqslant \Pr(d = c|\mathbf{x}, \boldsymbol{\nu})$ *if and only if* $W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) \geqslant W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}), \ (\mathbf{x}, \boldsymbol{\nu}) - a.s.$

Indeed, *any* model that satisfies an analogous property is incapable of explaining the relative frequency of \$200 in the distribution of observed deductible choices.[12] This includes, inter alia, the conditional logit model (McFadden 1974), the mixed logit model (McFadden 1974; McFadden and Train 2000), the multinomial probit model (e.g., Hausman and Wise 1978), and semiparametric models such as the one in Manski (1975). At the same time, not all choice set formation processes can explain the relative frequency of \$200 in our data. For instance, UR cannot but ASR can.

CLAIM S3.1: *Take the model in Section 2. Suppose for a given* $c \in \mathcal{D}$ *there exist* $a, b \in \mathcal{D}$*,* $a \neq b \neq c$*, such that for each* $\boldsymbol{\nu} \in \mathcal{V}$*,* $W(\mathbf{x}_a, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})$ *or* $W(\mathbf{x}_b, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})$*. Then for any distribution of* $\boldsymbol{\nu}$ *with support* $\mathcal{V}$*:*

(I) Property S3.1 implies $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) > \Pr(d = c|\mathbf{x}), \ \mathbf{x} - a.s.$

(II) Under UR, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) > \Pr(d = c|\mathbf{x}), \ \mathbf{x} - a.s.$

(III) Under ASR, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) < \Pr(d = c|\mathbf{x})$ *is possible.*

[12] In the case of a model with additively separable noise where $\boldsymbol{\nu} = (\boldsymbol{\upsilon}, (\epsilon_c, c \in \mathcal{D}))$ and $W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}) = \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta}) + \epsilon_c$, the analogous property is: For all $c, c' \in \mathcal{D}$, $\Pr(d = c'|\mathbf{x}, \boldsymbol{\upsilon}) \geqslant \Pr(d = c|\mathbf{x}, \boldsymbol{\upsilon})$ if and only if $\omega(\mathbf{x}_{c'}, \boldsymbol{\upsilon}; \boldsymbol{\delta}) \geqslant \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta}), \ (\mathbf{x}, \boldsymbol{\upsilon}) - a.s.$

Proof. The implication in Claim S3.1(I) follows from Property S3.1 by integrating with respect to the distribution of $\boldsymbol{\nu}$.

Claim S3.1(II) follows from the fact that the UR model satisfies Property S3.1. Suppose alternative c' is preferred to alternative c. Alternative c' may be chosen from choice sets that contain both c' and c and from choice sets that contain c' but not c. However, alternative c may be chosen only from choice sets that contain c but not c'. Because all choice sets, conditional on the draw of $|C|$, are equiprobable, c' is chosen more frequently than c.

We can establish Claim S3.1(III) with a trivial example. Suppose $\varphi(a) = \varphi(b) = 0$ and $\varphi(c) = 1$. Then $\Pr(d = a|\mathbf{x}) = \Pr(d = b|\mathbf{x}) = 0$ and $\Pr(d = c|\mathbf{x}) > 0$ provided there exists a positive measure of values $\boldsymbol{\nu} \in \mathcal{V}$ such that $W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta})$ for all $c' \in \mathcal{D}\backslash\{a, b\}$, $c' \neq c$. More generally, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) < \Pr(d = c|\mathbf{x})$ is possible provided $\varphi(a)$ and $\varphi(b)$ are sufficiently low, $\varphi(c)$ is sufficiently high, and c is the first best alternative in $\mathcal{D}\backslash\{a, b\}$ for some positive measure of values $\boldsymbol{\nu} \in \mathcal{V}$. □

We emphasize that Claim S3.1 does not rely on Assumption 3.1 or the assumptions of the empirical model in Section 4.1. It thus exemplifies a new approach to testing assumptions on choice set formation in any random utility model under weak restrictions on the utility function and without parametric restrictions on the distribution of preferences or choice sets.

An analogous claim holds in the case of a model with additively separable disturbances, such as the mixed logit model in Section 5.1.1, for any distribution of $\boldsymbol{\upsilon}$ with support Υ, where the predicate is: Suppose for a given $c \in \mathcal{D}$ there exist $a, b \in \mathcal{D}$, $a \neq b \neq c$, such that for each $\boldsymbol{\upsilon} \in \Upsilon$, $\omega(\mathbf{x}_a, \boldsymbol{\upsilon}; \boldsymbol{\delta}) > \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta})$ or $\omega(\mathbf{x}_b, \boldsymbol{\upsilon}; \boldsymbol{\delta}) > \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta})$.

References

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Barseghyan, L., F. Molinari, and M. Thirkettle (2020): "Discrete Choice under Risk with Limited Consideration," Working Paper, Department of Economics, Cornell University.

Barseghyan, L., J. C. Teitelbaum, and L. Xu (2018): "Different Contexts, Different Risk Preferences?" SSRN Working Paper No. 3200657.

Grant, M. and S. Boyd (2010): "CVX: Matlab Software for Disciplined Convex Programming, Version 1.21," available at http://cvxr.com/cvx.

Hausman, J. A. and D. A. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205 – 228.

McFadden, D. and K. Train (2000): "Mixed MNL Models for Discrete Reponse," *Journal of Applied Econometrics*, 15, 447–470.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

Molchanov, I. and F. Molinari (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.