cemmap

centre for microdata methods and practice

Heterogeneous choice sets and preferences

Levon Barseghyan
Maura Coughlin
Francesca Molinari
Joshua C. Teitelbaum

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP47/20

Economic and Social Research Council

Heterogeneous Choice Sets and Preferences*

Levon Barseghyan
Cornell University

Maura Coughlin
Rice University

Francesca Molinari
Cornell University

Joshua C. Teitelbaum
Georgetown University

September 21, 2020

Abstract

We propose a robust method of discrete choice analysis when agents' choice sets are unobserved. Our core model assumes nothing about agents' choice sets apart from their minimum size. Importantly, it leaves unrestricted the dependence, conditional on observables, between agents' choice sets and their preferences. We first characterize the sharp identification region of the model's parameters by a finite set of conditional moment inequalities. We then apply our theoretical findings to learn about households' risk preferences and choice sets from data on their deductible choices in auto collision insurance. We find that the data can be explained by expected utility theory with low levels of risk aversion and heterogeneous choice sets, and that more than three in four households require limited choice sets to explain their deductible choices. We also find that the data are consistent with some models of choice set formation, but not others.

Keywords: choice sets, discrete choice, partial identification, random utility, risk preferences, unobserved heterogeneity.

*We are grateful to the editor, the referees, Panle Jia Barwick, Aureo de Paula, Jean-Francois Houde, Chuck Manski, Ulrich Müller, Matthew Thirkettle, Lin Xu, and seminar and conference participants at the 2016 Conference of the International Association for Applied Econometrics, the 2016 Penn State-Cornell Conference on Econometrics and IO, the 2016 Boston College-UCL Conference on Heterogeneity in Supply and Demand, the 2017 ASSA Meetings, the 2018 Conference of Former Northwestern Econometrics Students, the 2018 Northwestern-CeMMAP Conference on Incomplete Models, the 2019 California Econometrics Conference, the 2019 Armenian Economic Association Meetings, the 2019 EGSC, the 2020 ESWC, Bocconi, BU, Bristol, Caltech, Cornell, Duke, JHU, Leuven, Microsoft, Penn, Princeton, Rice, Tilbergen, UCL, UNC, USC, UVA, Warwick, and Washington. We acknowledge financial support from National Science Foundation grants SES-1031136 and SES-1824448 and from the Institute for Social Sciences at Cornell University. Part of the research for this paper was conducted while Barseghyan and Molinari were on sabbatical leave at the Department of Economics at Duke University, whose hospitality is gratefully acknowledged.

1 Introduction

The starting point of any discrete choice problem is the finite set of alternatives from which the agent makes her choice—her choice set. Discrete choice analysis in the tradition of McFadden (1974) rests on two assumptions about agents' choice sets. The first is that an agent's choice set is a subset of a known universal set of feasible alternatives—the feasible set. The second assumption is that an agent's choice set is observed. McFadden shows that when these assumptions hold, one can apply the principle of revealed preference to learn about agents' unobserved preferences from data on their observed choices. Moreover, he shows that with additional restrictions on the structure and distribution of agents' preferences, one can achieve point identification of a parametric model of discrete choice.

In practice, however, agents' choice sets are often unobserved. Sometimes this is a missing data problem—the agents' choice sets are observable in principle but are not recorded in the data. For example, one studying the college enrollment choices of high school students may not observe the colleges to which a student applied and was admitted (Kohn et al. 1976); one studying the travel mode choices of urban commuters may not observe if some modes normally available to a commuter were temporarily unavailable on a given day (Ben-Akiva and Boccara 1995); or one studying the hospital choices of English patients may not observe which alternatives were offered to a patient by her referring physician (Gaynor et al. 2016).

At other times the problem is that agents' choice sets are unobservable mental constructs. This is the case in models of limited attention or limited consideration, where an agent considers only a subset of the feasible set due to, for example, search costs, brand preferences, or cognitive limitations. For instance, one studying the personal computer choices of retail consumers can be sure that a consumer was not aware of all computers for sale but cannot observe the computers of which a consumer was aware (Goeree 2008); one studying the Medigap plan choices of Medicare insureds cannot observe which of the available plans an insured in fact considered (Starc 2014); or one studying the energy retailer choices of residential electricity customers cannot observe whether or to what extent a customer considered the alternatives to her default, incumbent retailer (Hortaçsu et al. 2017).

When agents' choice sets are unobserved the econometrician is forced to make additional assumptions in order to achieve point identification. The most common approach is to assume, often implicitly, that all choice sets coincide with the feasible set or a known subset of the feasible set. More sophisticated approaches allow for heterogeneity in agents' choice sets and obtain point identification by relying on auxiliary information about the composition or distribution of choice sets, two-way exclusion restrictions (i.e., variables assumed to impact choice sets but not preferences and vice versa), and other restrictions on the choice set formation process (e.g., conditional independence between choice sets and preferences). In some applications these approaches seem reasonable or at least plausible. In many applications, however, they likely result in misspecified models, biased estimates, and incorrect inferences.

More fundamentally, the basic revealed preference argument breaks down when choice sets are unobserved. At one extreme, when an agent's choice set equals the feasible set, her choice reveals that she prefers the chosen alternative to all others. At the other extreme, when an agent's choice set comprises a single alternative, her choice is driven entirely by her choice set and reveals nothing about her preferences. In all other cases her choice is a function of both her preferences and her choice set. Learning about preferences from choices when choice sets are unobserved is the main challenge we address in this paper.

We propose a new, robust method of discrete choice analysis when agents' choice sets are unobserved. Our core model imposes mild restrictions on agents' preferences and assumes nothing about agents' choice sets or how they are formed, apart from assuming that they have a known minimum size greater than one. Under these weak conditions the distribution of preferences typically is not point identified, but only partially identified. In our main theoretical result we characterize the sharp identification region of the distribution of preferences. Sharpness means that the identification region comprises all and only those preference distributions for which there exists a choice set formation process such that the distribution of model implied choices matches the distribution of observed choices. As a corollary to our main result we show that if one also assumes that preferences are independent of choice set size then the distribution of choice set size is also partially identified.

We lay out our core model in Section 2. We begin with the classic random utility model developed by McFadden (1974) and others, though we allow for a utility function that is neither linear in parameters nor additively separable in unobservables. Our key point of departure from the classic model, however, is that we relax the assumption that the agents' choice sets are observed. Instead, we assume only that the minimum size of the agents' choice sets is a known integer greater than one. Consequently, our model admits any choice set formation process (subject only to the minimum size assumption) and allows for any dependence structure, without restriction, between agents' choice sets and their observables and, conditional on observables, between agents' choice sets and their preferences.

In Section 3 we first show that our model implies multiple optimal choices for an agent, resulting from the multiple possible realizations of her choice set. It is this multiplicity that, in the absence of additional restrictions on the choice set formation process, generally precludes point identification of the model's parameters. Because we avoid making such additional, unverifiable assumptions, our approach yields a robust method of statistical inference. We then present our main identification results, both of which leverage a result in random set theory, due to Artstein (1983), to define a finite set of conditional moment inequalities that characterizes the sharp identification region for the model's parameters.[1]

In Sections 4 and 5 we demonstrate the usefulness of our theoretical findings by applying them to learn about households' risk preferences and choice sets from data on their deductible choices in auto collision insurance. The data hail from a large U.S. insurance company and contain information on more than 100,000 households who first purchased auto collision coverage from the company between 1998 and 2007.

In Section 4 we specify an empirical model of deductible choice in auto collision insurance that allows for unobserved heterogeneity in households' risk preferences and in their choice sets. Although we observe the feasible set of deductibles, we do not observe which deductibles enter a household's choice set. In our setting, therefore, there may be unobserved heterogeneity in choice sets. What's more, such unobserved heterogeneity may be due to missing data—e.g., if different households are quoted different subsets of deductibles—or to unobserved constraints—e.g., if some households disregard low deductibles due to budget constraints or high deductibles due to liquidity constraints. Either way, the robust method of discrete choice analysis that we propose is equally applicable and valid.

We present our empirical findings in Section 5. Our key finding with respect to risk aversion is that the data can be explained by expected utility theory with a distribution of risk aversion that has low mean and variance, with at least a quarter of households being effectively risk neutral. By comparison, three point identified expected utility models—two that specify processes of heterogeneous choice set formation and are estimated on our data, and one that assumes full size choice sets (i.e., choice sets that contain all feasible alternatives) and is estimated on similar data—yield means that are substantially higher.

Our key finding with respect to choice sets is that more than three in four households require limited (i.e., less than full size) choice sets to explain their deductible choices. We discuss two drivers of this result—suboptimal choices and violations of the law of demand. We also discuss how the frequency of suboptimal choices is consistent with some models of choice set formation, but not others. The latter discussion contributes a new, robust approach to testing the assumptions on choice set formation in a random utility model.

Our empirical findings highlight the importance of using a robust method to conduct inference on discrete choice models when there may be unobserved heterogeneity in choice sets. The literature on risky choice, motivated in part by reported estimates of risk aversion that seem implausibly high in light of the Rabin (2000) critique (e.g., Cicchetti and Dubin

[1]The recent econometrics literature uses the result in Artstein (1983), discussed in detail in Molchanov and Molinari (2018, Chapter 2), to conduct identification analysis in various partially identified models (e.g., Beresteanu and Molinari 2008; Beresteanu et al. 2011; Galichon and Henry 2011; Chesher et al. 2013; Chesher and Rosen 2017). For a review, see Molinari (2020).

1994; Sydnor 2010), has focused on developing and estimating models that depart from expected utility theory in their specification of *how* agents evaluate risky alternatives. Our findings provide new evidence on the importance of developing models that differ in their specification of *which* alternatives agents evaluate, and of data collection efforts that seek to directly measure agents' heterogeneous choice sets (Caplin 2016).

We conclude the paper in Section 6 with a discussion in which we provide an overview of the prior literature on discrete choice analysis with unobserved heterogeneity in choice sets and recap our contributions to the literature.

2 A Random Utility Model with Unobserved Heterogeneity in Choice Sets

Our starting point is the random utility model developed by McFadden (1974). Let $\mathcal{I}$ denote a population of agents and $\mathcal{D}$ denote a finite set of alternatives, which we call the *feasible set*. Let $\mathcal{U}$ be a family of real valued functions defined over the elements of $\mathcal{D}$. The random utility model posits that for each agent $i \in \mathcal{I}$ there exists a function U_i drawn from $\mathcal{U}$ according to some probability distribution such that

$$d \in^* C_i \;\Leftrightarrow\; U_i(d) \geqslant U_i(c) \text{ for all } c \in C_i,\; c \neq d, \tag{2.1}$$

where $\in^*$ denotes "is chosen from" and $C_i \subseteq \mathcal{D}$ denotes the agent's *choice set*.

We assume that each agent $i \in \mathcal{I}$ is characterized by a real valued vector of observable attributes $\mathbf{x}_i = (\mathbf{s}_i, (\mathbf{z}_{ic}, c \in \mathcal{D}))$, where $\mathbf{s}_i$ is a subvector of attributes specific to agent i that are constant across alternatives and $\mathbf{z}_{ic}$ is a subvector of attributes specific to alternative c that may vary across agents. Let $\mathbf{x}_{ic} = (\mathbf{s}_i, \mathbf{z}_{ic})$ denote the vector of observable attributes relevant to alternative c. In addition, we assume that each agent $i \in \mathcal{I}$ is further characterized by a real valued vector of unobservable attributes $\boldsymbol{\nu}_i$, which are idiosyncratic to the agent. Let $\mathcal{X}$ and $\mathcal{V}$ denote the supports of $\mathbf{x}_i$ and $\boldsymbol{\nu}_i$, respectively.

To operationalize U_i as a random variable, we posit that it is a function of the agent's observable and unobservable attributes and we impose restrictions on its distribution.

Assumption 2.1 (Restrictions on Utility):

(I) There exists a function $W : \mathcal{X} \times \mathcal{V} \mapsto \mathbb{R}$, known up to a finite dimensional parameter vector $\boldsymbol{\delta} \in \Delta \subset \mathbb{R}^k$, where Δ is convex and compact, and continuous in each of its arguments such that $U_i(c) = W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ for all $c \in \mathcal{D}$, $(\mathbf{x}_{ic}, \boldsymbol{\nu}_i) - a.s.$

(II) The distribution of $\boldsymbol{\nu}_i$, denoted by P, is continuous, known up to a finite dimensional parameter vector $\boldsymbol{\gamma} \in \Gamma \subset \mathbb{R}^l$, where Γ is convex and compact, and independent of $\mathbf{x}_i$.

Assumption 2.1 allows for nonadditive unobserved heterogeneity in U_i, indexed by $\boldsymbol{\nu}_i$. It is weaker than the standard assumption that U_i is additively separable in unobservables. That said, one could let $\boldsymbol{\nu}_i = (\nu_{ic}, c \in \mathcal{D})$ and specify $W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \omega(\mathbf{x}_{ic}; \boldsymbol{\delta}) + \nu_{ic}$ as in a conditional logit (McFadden 1974), or let $\boldsymbol{\nu}_i = (\boldsymbol{\upsilon}_i, (\epsilon_{ic}, c \in \mathcal{D}))$ and specify $W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \omega(\mathbf{x}_{ic}, \boldsymbol{\upsilon}_i; \boldsymbol{\delta}) + \epsilon_{ic}$ as in a mixed logit (McFadden and Train 2000).

Assumption 2.1 also posits that the functional family of U_i and the distributional family of $\boldsymbol{\nu}_i$ are known parametric classes, and that $\boldsymbol{\nu}_i$ is independent of $\mathbf{x}_i$. Though standard in discrete choice analysis, the parametric assumptions are not essential for our partial identification results (see Remark 3.1), and the independence assumption can be relaxed based on the specific structure of the empirical model (as we illustrate in our application). The assumption that P is continuous, which ensures there are no utility ties, is also nonessential because our partial identification results allow for sets of model implied optimal choices and thus can readily accommodate utility ties; see Section A.3.3 of the Supplemental Material.

Our key point of departure from McFadden (1974) and the bulk of the discrete choice literature is the assumption regarding what is observed by the econometrician. It is standard to assume that (i) a random sample of choice sets C_i, choices d_i, and attributes $\mathbf{x}_i$, $\{(C_i, d_i, \mathbf{x}_i) : d_i \in^* C_i, i \in I \subset \mathcal{I}\}$, is observed, and that (ii) $|C_i| \geqslant 2$ for all $i \in \mathcal{I}$, where $|\cdot|$ denotes set cardinality (see, e.g., Manski 1975, Assumption 1). By contrast, we assume:

Assumption 2.2 (Random Sample and Minimum Choice Set Size):

(I) A random sample of choices d_i and attributes $\mathbf{x}_i$, $\{(d_i, \mathbf{x}_i) : i \in I \subset \mathcal{I}\}$, is observed.

(II) $\Pr(|C_i| \geqslant \kappa) = 1$ for all $i \in \mathcal{I}$, where $\kappa \geqslant 2$ is a known integer.

Assumption 2.2(I) is weaker than the standard assumption as it omits the requirement that choice sets are observed. Given this difference, Assumption 2.2(II) is comparable to the standard assumption as it requires that choice sets have a known minimum size, κ, greater than one. The empirical content of the model increases with κ. Knowledge of κ is immediate when choice sets are observed. We assume that κ is known, either from information in the data or by assumption, even though choice sets are unobserved. In any event, Assumption 2.2(II) is weaker than the assumption that every agent's choice set coincides with the feasible set or a known subset of the feasible set.

Remark 2.1: Assumption 2.2(II) can be weakened to $\Pr(|C_i| = 1) \leqslant \bar{\pi}_1 < 1$ for all $i \in \mathcal{I}$, where $\bar{\pi}_1$ is known. The empirical content of the model is decreasing in $\bar{\pi}_1$.

A key feature of our model is that it admits *any* choice set formation process, including any mixture process, subject only to Assumption 2.2(II). Choice sets may be formed by internal processes, such as simultaneous or sequential search (Stigler 1961; Weitzman 1979; Honka et al. 2019) or elimination-by-aspects or attention or attribute filters (Tversky 1972a,b; Masatlioglu et al. 2012; Kimya 2018; Cattaneo et al. 2020), or by external processes, such as advertising (Chamberlin 1933; Goeree 2008; Terui et al. 2011) or choice architecture (Thaler and Sunstein 2008; Johnson et al. 2012; Gaynor et al. 2016). Whether internal or external, the choice set formation process can admit any dependence structure, without restriction, between agents' choice sets and their observable attributes and, conditional on observables, between agents' choice sets and their unobservable attributes. That is, C_i can be arbitrarily correlated with $\mathbf{x}_i$ and, conditional on $\mathbf{x}_i$, C_i can be arbitrarily correlated with $\boldsymbol{\nu}_i$.

3 Partial Identification of the Model's Parameters

3.1 Preferences

The random utility model in Section 2 implies multiple optimal choices for the agent, due to the multiple possible realizations G of her choice set C_i. Let $d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ denote the model implied optimal choice for agent i with attributes $(\mathbf{x}_i, \boldsymbol{\nu}_i)$, choice set $C_i = G \subseteq \mathcal{D}$, $|G| \geqslant \kappa$, and utility parameter $\boldsymbol{\delta}$. That is, $d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \equiv \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$.

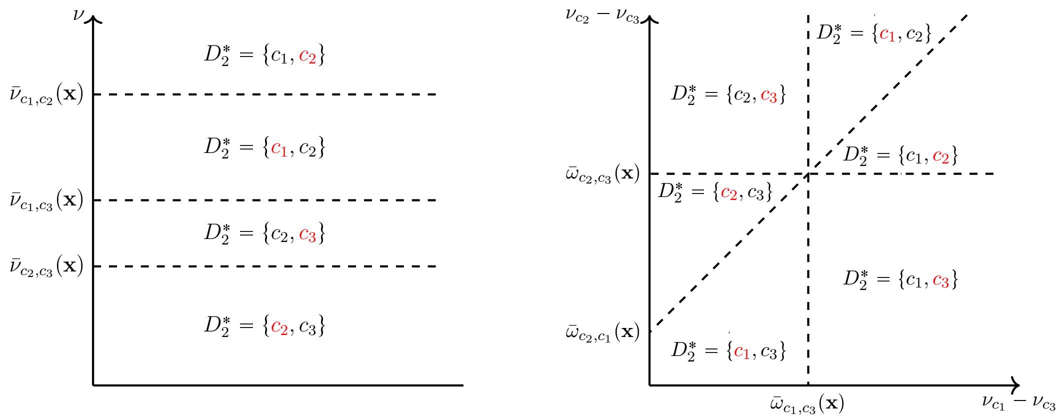
The set of model implied optimal choices given $(\mathbf{x}_i, \boldsymbol{\nu}_i)$ and $\boldsymbol{\delta}$ is

$$D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \bigcup_{G \subseteq \mathcal{D}: |G| \geqslant \kappa} \left\{ d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \left\{ d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\}, \tag{3.1}$$

where the last equality follows from Sen's property α: any alternative that is optimal for a given choice set $G' \subseteq \mathcal{D}$ is also optimal for every choice set $G \subset G'$ containing that alternative. The set $D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is a *random closed set* with realizations in $\mathcal{D}$.[2] It contains the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, where "best" is defined with respect to U_i. Figure 3.1 contains stylized depictions of $D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ when $|\mathcal{D}| = 3$ and $\kappa = 2$.

When the information in the data and the economic model do not impose sufficiently strong restrictions on the distribution of C_i, the multiplicity of model implied optimal choices generally precludes point identification of the model's parameters $\boldsymbol{\theta} = [\boldsymbol{\delta}; \gamma]$. The reason is that the relationship between the model and the data is *incomplete* (Tamer 2003). To see this, let $\Pr(d_i^* = c | \mathbf{x}_i; \boldsymbol{\theta}, \mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i))$ denote the model implied conditional probability that alternative c is chosen given $\mathbf{x}_i$ and $(\boldsymbol{\theta}, \mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i))$, where $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ denotes the conditional

[2]The formal definition of a random closed set is provided in Definition A.1 in the Supplemental Material. That $D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is a random closed set is formally established in Lemma A.1 in the Supplemental Material.



(a) $U(c) = W(\mathbf{x}_c, \nu; \boldsymbol{\delta})$ (b) $U(c) = \omega(\mathbf{x}_c; \boldsymbol{\delta}) + \nu_c$

Figure 3.1: Stylized depictions of D_κ^* when $|\mathcal{D}| = 3$ and $\kappa = 2$.

Notes: In Panel (a), $\boldsymbol{\nu} \in \mathbb{R}$, $U(c) = W(\mathbf{x}_c, \nu; \boldsymbol{\delta})$, and the alternatives in $\mathcal{D}$ are vertically differentiated. The threshold $\bar{\nu}_{c_a,c_b}(\mathbf{x})$ is the value of ν above which c_a has greater utility than c_b and below which c_b has greater utility than c_a. In Panel (b), $\boldsymbol{\nu} \in \mathbb{R}^3$ and $U(c) = \omega(\mathbf{x}_c; \boldsymbol{\delta}) + \nu_c$. The threshold $\bar{\omega}_{c_a,c_b}(\mathbf{x}) \equiv \omega(\mathbf{x}_{c_b}; \boldsymbol{\delta}) - \omega(\mathbf{x}_{c_a}; \boldsymbol{\delta})$ is the value of $\nu_{c_a} - \nu_{c_b}$ above which c_a has greater utility than c_b and below which c_b has greater utility than c_a. Because $\kappa = 2$, either $|C| = 2$ or $|C| = 3$ and hence D_2^* comprises the first and second best alternatives in $\mathcal{D}$. For a given $\boldsymbol{\nu}$, the first best appears in black and the second best in red. The agent's choice is determined by her realization G of C. She chooses the first best if it is in G; otherwise she chooses the second best.

probability mass function of C_i given $(\mathbf{x}_i, \boldsymbol{\nu}_i)$. For all $c \in \mathcal{D}$,

$$\Pr(d_i^* = c|\mathbf{x}_i; \boldsymbol{\theta}, \mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)) = \int_{\boldsymbol{\tau} \in \mathcal{V}} \sum_{G \subseteq \mathcal{D}} \mathbf{1}(d_i^*(G; \mathbf{x}_i, \boldsymbol{\tau}; \boldsymbol{\delta}) = c) \mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\tau}) dP(\boldsymbol{\tau}; \boldsymbol{\gamma}). \tag{3.2}$$

Because we require only that $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = 0$ for $G \subseteq \mathcal{D}$, $|G| < \kappa$, there may be multiple admissible values of $(\boldsymbol{\theta}, \mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i))$ such that

$$\Pr(d_i^* = c|\mathbf{x}_i; \boldsymbol{\theta}, \mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)) = \Pr(d_i = c|\mathbf{x}_i), \ \forall c \in \mathcal{D}, \ \mathbf{x}_i - a.s., \tag{3.3}$$

where d_i is the agent's observed choice.[3] Nonetheless, in general, it is not the case that for every $\boldsymbol{\theta}$ in a parameter space Θ there is an admissible $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ such that condition (3.3) holds. Hence, we can partially identify $\boldsymbol{\theta}$ from the information in the data and the model.

The set of values of $\boldsymbol{\theta} \in \Theta$ for which there exists an admissible distribution $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ such that condition (3.3) holds forms the *sharp identification region* for $\boldsymbol{\theta}$. We denote this region by Θ_I. The distribution $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$, however, is an infinite dimensional nuisance

[3]By contrast, if $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ is known or sufficiently restricted (e.g., parametrically specified), then $\boldsymbol{\theta}$ can be point identified by condition (3.3) given sufficient variation in $\mathbf{x}_i$.

parameter, which creates difficulties for the computation of Θ_I and for statistical inference. We circumvent these difficulties by working directly with the set $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$.

If the model is correctly specified, the agent's observed choice d_i is maximal with respect to her preference among the alternatives in her choice set and it therefore satisfies

$$d_i \in D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}),\ \ \mathbf{x}_i - a.s., \tag{3.4}$$

for the data generating value of $\boldsymbol{\theta}$. To harness the empirical content of equation (3.4), we leverage a result in Artstein (1983), reported in Theorem A.1 in the Supplemental Material. This result allows us to translate equation (3.4) into a finite number of conditional moment inequalities that fully characterize the sharp identification region Θ_I.

Theorem 3.1: *Let Assumptions 2.1 and 2.2 hold. In addition, let* $\boldsymbol{\theta} = [\boldsymbol{\delta}; \boldsymbol{\gamma}]$, $\Theta = \Delta \times \Gamma$, *and* $\mathbb{K} = \{K \subseteq \mathcal{D} : |K| < \kappa\}$. *Then*

$$\Theta_I = \left\{ \boldsymbol{\theta} \in \Theta : \Pr(d \in K|\mathbf{x}) \leqslant P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \forall K \in \mathbb{K}, \mathbf{x} - a.s. \right\}. \tag{3.5}$$

Our proof of Theorem 3.1, provided in Section A.3.1 of the Supplemental Material, establishes that the characterization in equation (3.5) is sharp—all and only those values of $\boldsymbol{\theta} \in \Theta$ for which the inequalities in equation (3.5) hold could have generated the observed data under the maintained assumptions.[4] These inequalities have a straightforward interpretation. At the data generating value of $\boldsymbol{\theta} \in \Theta$, it must be the case that, for every subset $K \in \mathbb{K}$, the conditional probability that K contains a model implied optimal choice (right hand side) is not less than the conditional probability of the observed choice (left hand side), which itself is optimal. When the alternatives in $\mathcal{D}$ are vertically differentiated, the set $\mathbb{K}$ can be restricted to the subsets $\overrightarrow{K} = \{c_1\}, \{c_1, c_2\}, \ldots, \{c_1, c_2, \ldots, c_{\kappa-1}\}$ and $\overleftarrow{K} = \{c_{|\mathcal{D}|}\}, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}\}, \ldots, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}, \ldots, c_{|\mathcal{D}|-\kappa+2}\}$,[5] and the inequalities translate into statements about cumulative shares for higher (respectively lower) quality alternatives.

3.1.1 Computational Feasibility

There are two main computational challenges in applying Theorem 3.1. First, given any $\kappa \geqslant 2$, the number of inequalities in equation (3.5) grows superlinearly with $|\mathcal{D}|$. In Theorem A.2 and Corollary A.2 in the Supplemental Material, we provide sufficient conditions to reduce the number of inequalities needed to obtain Θ_I. Exploiting these results, in simulations (available upon request) we run our analysis with $|\mathcal{D}| = 101$, $\kappa = 10, 30, 50, 70, 90$, and

[4] If per Remark 2.1 one weakens Assumption 2.2(II) to $\Pr(|C_i| = 1) \leqslant \bar{\pi}_1 < 1$ where $\bar{\pi}_1$ is known, then $\Theta_I = \{\boldsymbol{\theta} \in \Theta : \Pr(d \in K|\mathbf{x}) \leqslant \bar{\pi}_1 + (1 - \bar{\pi}_1) P(D^*_2(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \forall K \in \mathbb{K}, \mathbf{x} - a.s.\}$.

[5] See Corollary A.2 in the Supplemental Material.

$\boldsymbol{\nu} \in \mathbb{R}$. In each case we compute a 95 percent confidence set for Θ_I,[6] with the number of inequalities ranging from about 1,800 to 18,000, in less than half an hour. Moreover, in Corollary A.1, we show that Θ_I can be equivalently characterized as the set of values $\boldsymbol{\theta} \in \Theta$ for which the optimal value of a convex program is zero. Because one can leverage efficient algorithms for solving convex programs, the computational cost of the convex programming method grows with $|\mathcal{D}|$ at a slower rate than the cost of the direct method.

The second challenge is computing the model implied probabilities (the right hand sides of the inequalities).[7] In Theorem A.3 in the Supplemental Material, we provide substantial simplifications to compute these probabilities when the dimension of $\boldsymbol{\nu}$ is large. We do so in mixed logit models with unobserved heterogeneity in choice sets, where choice sets can be arbitrarily correlated with the entire vector $\boldsymbol{\nu}$, but the random coefficients and the additive disturbances are independent. We show how one can exploit the logit closed form choice probabilities and then numerically integrate over the random coefficients. We illustrate this in simulations (available upon request) where $|\mathcal{D}| = 7, 12, 17$, $\kappa = 5$, and $\boldsymbol{\nu} \in \mathbb{R}^{|\mathcal{D}|+1}$. The choices of $|\mathcal{D}|$ and κ are motivated by recent studies of drug plan choices under limited consideration in Medicare Part D (Abaluck and Adams 2020; Coughlin 2020). In each case we compute a 95 percent confidence set for Θ_I, with the number of inequalities ranging from about 10,000 to 325,000, and the run times ranging from about 5 minutes to 20 hours.

To evaluate the feasibility of our method from the perspective of a researcher who has access to run-of-the-mill computing power (as opposed to large clusters or cloud computing), we run our simulations on a single, four year old Dell Precision Tower 7910 (Dual CPU E5-2687W v4 @ 3.00GHz with 128GM memory).

3.2 Choice Sets

Theorem 3.1 establishes that, under mild restrictions on the utility function (Assumption 2.1) and knowing only the minimum size of agents' choice sets (Assumption 2.2), one can learn features of the distribution of preferences without observing agents' choice sets or knowing how they are formed. We now show that, with an additional restriction on the choice set formation process, one can also learn features of the distribution of choice sets.

Let $\ell_i \equiv |C_i|$ denote the size of agent i's choice set C_i. When $\ell_i = |\mathcal{D}|$ we say that C_i has "full" size. When $\ell_i < |\mathcal{D}|$ we say that C_i is "limited" or "restricted." More specifically, we say that C_i is "full-1" when $\ell_i = |\mathcal{D}| - 1$, "full$-2$" when $\ell_i = |\mathcal{D}| - 2$, and so forth.

In addition to Assumptions 2.1 and 2.2, one could assume that:

[6] See Section 5.

[7] The left hand side of each inequality can be estimated from the data. The right hand side is a function of $\mathbf{x}_i$ known up to $\boldsymbol{\theta}$.

Assumption 3.1 (Choice Set Size): *Agent i draws the size ℓ_i of her choice set such that*

$$\Pr(\ell_i = q|\mathbf{x}_i, \boldsymbol{\nu}_i) = \Pr(\ell_i = q|\mathbf{x}_i) = \pi(q; \mathbf{x}_i; \boldsymbol{\eta}), \ q = \kappa, \ldots, |\mathcal{D}|, \tag{3.6}$$

where $\pi(q; \mathbf{x}_i; \boldsymbol{\eta}) \geqslant 0$ for $q \geqslant \kappa$, $\sum_{q=\kappa}^{|\mathcal{D}|} \pi(q; \mathbf{x}_i; \boldsymbol{\eta}) = 1$, and the function π is known up to a finite dimensional parameter vector $\boldsymbol{\eta} \in H \subset \mathbb{R}^m$ where H is convex and compact. To simplify notation, define $\pi_q(\mathbf{x}; \boldsymbol{\eta}) \equiv \pi(q; \mathbf{x}; \boldsymbol{\eta})$.

Assumption 3.1 posits that the size ℓ_i of agent i's choice set is drawn from an unspecified distribution with support $\{\kappa, \ldots, |\mathcal{D}|\}$, which allows for the possibility that the agent's choice set has full size, $\ell_i = |\mathcal{D}|$, or is limited, $\ell_i < |\mathcal{D}|$. The only restrictions it imposes on the distribution of agents' choice sets are that the distributional family of ℓ_i is a known parametric class—though, as before, the parametric structure is not essential (see Remark 3.1)—and that ℓ_i is independent of $\boldsymbol{\nu}_i$. Conditional on ℓ_i, however, the model with Assumption 3.1 continues to allow for any dependence structure, without restriction, between agents' choice sets and their observable attributes and, conditional on observables, between agents' choice sets and their unobservable attributes. Moreover, agents with choice sets of the same size need not have choice sets with the same composition.

Under Assumption 3.1, Theorem 3.1 specializes to the following corollary.[8]

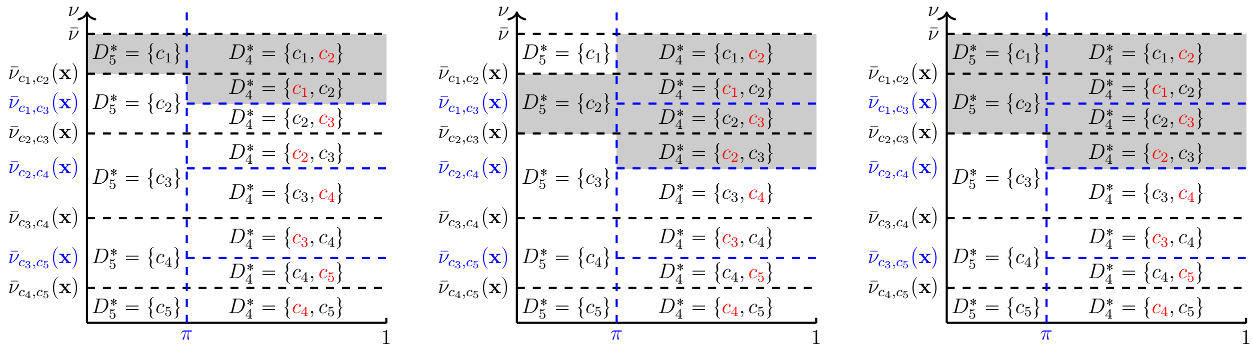
Corollary 3.1: *Let Assumptions 2.1, 2.2, and 3.1 hold. In addition, let $\boldsymbol{\theta} = [\boldsymbol{\eta}; \boldsymbol{\delta}; \boldsymbol{\gamma}]$ and $\Theta = H \times \Delta \times \Gamma$. Then*

$$\Theta_I = \left\{ \boldsymbol{\theta} \in \Theta : \Pr(d \in K|\mathbf{x}) \leqslant \textstyle\sum_{q=\kappa}^{|\mathcal{D}|} \pi_q(\mathbf{x}; \boldsymbol{\eta}) P(D_q^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \forall K \subseteq \mathcal{D}, \mathbf{x} - a.s. \right\}. \tag{3.7}$$

The sharp identification region Θ_I in Corollary 3.1 has two noteworthy features. First, the projection of Θ_I on $[\boldsymbol{\delta}; \boldsymbol{\gamma}]$ is equal to the sharp identification region in Theorem 3.1. In other words, the information in Θ_I about the distribution of preferences is the same with or without Assumption 3.1. This is because $D_{q+1}^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \subset D_q^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ for all $q \geqslant \kappa$, and thus the projection of Θ_I on $[\boldsymbol{\delta}; \boldsymbol{\gamma}]$ is obtained with $\pi_\kappa(\mathbf{x}_i; \boldsymbol{\eta}) = 1$ and $\pi_q(\mathbf{x}_i; \boldsymbol{\eta}) = 0$ for $q > \kappa$. Second, Θ_I provides information about the distribution of choice set size, as well. It yields a lower bound on $\pi_\kappa(\mathbf{x}_i; \boldsymbol{\eta})$ (the upper bound is one provided $\kappa < |\mathcal{D}|$) and upper bounds on $\pi_q(\mathbf{x}_i; \boldsymbol{\eta})$ for $q = \kappa + 1, \ldots, |\mathcal{D}|$ (the lower bounds are zero provided $\kappa < |\mathcal{D}|$ because $D_{q+1}^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \subset D_q^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$).[9]

[8]The proof of Corollary 3.1 follows immediately from the proof of Theorem 3.1 and therefore is omitted.

[9]The computational burden of recovering Θ_I under Corollary 3.1 may be greater than under Theorem 3.1 because more inequalities may be needed. However, one can still take advantage of the strategies referenced in Section 3.1.1 to reduce the computational burden.



(a) $\Pr(d = c_1|\mathbf{x}) \leqslant \Pr(\text{shaded area})$ (b) $\Pr(d = c_2|\mathbf{x}) \leqslant \Pr(\text{shaded area})$ (c) $\Pr(d \in \{c_1, c_2\}|\mathbf{x}) \leqslant \Pr(\text{shaded area})$

Figure 3.2: Stylized depictions of inequalities in Θ_I when $|\mathcal{D}| = 5$ and $\kappa = 4$.

Notes: Inequalities for three sets $K \subseteq \mathcal{D}$ are depicted: (a) $K = \{c_1\}$; (b) $K = \{c_2\}$; and (c) $K = \{c_1, c_2\}$. For a given $\boldsymbol{\nu}$, the first best alternative in $\mathcal{D}$ appears in black and the second best in red.

Remark 3.1: Theorem 3.1 and Corollary 3.1 can be generalized for a *structure* (W, P) or (W, P, π), as the case may be, that is subject only to nonparametric restrictions. We focus on the case with parametric restrictions for computational reasons and because methods of statistical inference for moment inequality models focus on this case.

3.3 Illustration of the Inequalities Characterizing Θ_I

Figure 3.2 contains stylized depictions of three inequalities in equation (3.7) when $|\mathcal{D}| = 5$, $\kappa = 4$, $\boldsymbol{\nu}_i = \nu_i$ is a scalar with support $\mathcal{V} = [0, \bar{\nu}]$, and the alternatives in $\mathcal{D}$ are vertically differentiated. In this case $\Pr(\ell_i \in \{4, 5\}) = 1$, and with a slight abuse of notation we let $\pi = \Pr(\ell_i = 5|\mathbf{x}_i)$. Thus, with probability π the agent draws a choice set of size 5, in which case D_5^* comprises the first best alternative. With probability $1 - \pi$ the agent draws a choice set of size 4, in which case D_4^* comprises the first and second best alternatives. In the former case the agent chooses the first best alternative. In the latter case the agent's choice is determined by her realization G of C_i. She chooses the first best if it is contained in G; otherwise she chooses the second best.[10] The threshold $\bar{\nu}_{c_a,c_b}(\mathbf{x}_i)$ is the value of ν_i above which c_a has a greater utility than c_b and below which c_b has a greater utility than c_a.

Panel (a) depicts the inequality for $K = \{c_1\}$. If $\ell_i = 5$ then $C_i = \mathcal{D}$ and c_1 is the optimal choice if $\nu_i > \bar{\nu}_{c_1,c_2}(\mathbf{x}_i)$. If $\ell_i = 4$ then c_1 is optimal if $\nu_i > \bar{\nu}_{c_1,c_2}(\mathbf{x}_i)$ and the realization G of

[10]In general, the agent chooses the best alternative in the intersection of her realizations of D_q^* and C_i.

C_i includes c_1 or if $\nu_i \in [(\bar{\nu}_{c_1,c_3}(\mathbf{x}_i), \bar{\nu}_{c_1,c_2}(\mathbf{x}_i)]$ and G excludes c_2. It follows that

$$\Pr(d_i = c_1|\mathbf{x}_i) \leqslant \pi P(\nu_i > \bar{\nu}_{c_1,c_2}(\mathbf{x}_i); \boldsymbol{\gamma}) + (1-\pi) P(\nu_i > \bar{\nu}_{c_1,c_3}(\mathbf{x}_i); \boldsymbol{\gamma}).$$

Similar reasoning applies to the other singleton sets, with $K = \{c_2\}$ depicted in Panel (b).

The inequalities in equation (3.7) also include those for non-singleton sets. To see why, Panel (c) depicts the inequality for $K = \{c_1, c_2\}$. While the left hand side is additive,

$$\Pr(d_i \in \{c_1, c_2\}|\mathbf{x}_i) = \Pr(d_i = c_1|\mathbf{x}_i) + \Pr(d_i = c_2|\mathbf{x}_i),$$

the right hand side is subadditive: the shaded area in Panel (c) is smaller than the sum of the shaded areas in Panels (a) and (b). Hence, values of $\boldsymbol{\theta} \in \Theta$ that satisfy the inequalities for $K = \{c_1\}$ and $K = \{c_2\}$ may fail to satisfy the inequality for $K = \{c_1, c_2\}$.

Not all pairs of singleton sets, however, yield nonredundant inequalities. Consider, for example, $K = \{c_1\}$ and $K = \{c_5\}$. As is clear from Figure 3.2, there is no value of ν_i for which D_4^* contains both c_1 and c_5. It follows that the inequality for $K = \{c_1, c_5\}$ is redundant if the inequalities for $K = \{c_1\}$ and $K = \{c_5\}$ are satisfied. This reasoning can substantially reduce the number of inequalities needed to recover Θ_I; see Section 3.1.1.

Though not depicted in Figure 3.2, let us highlight the algebra that delivers an upper bound on π. Consider $K = \{c_1, c_2, c_3, c_4\}$. Given this K we have

$$\begin{aligned} &\Pr(d_i \in K|\mathbf{x}_i) \leqslant \pi \Pr(D_5^* \cap K \neq \varnothing) + (1-\pi) \Pr(D_4^* \cap K \neq \varnothing) \\ &\qquad \Leftrightarrow \quad \Pr(d_i = c_5|\mathbf{x}_i) \geqslant \pi \Pr(D_5^* = \{c_5\}) = \pi P(\nu_i \leqslant \bar{\nu}_{c_4,c_5}(\mathbf{x}_i); \boldsymbol{\gamma}). \end{aligned}$$

Given any $\boldsymbol{\gamma}$, this inequality yields an upper bound on π. In general, one obtains the upper bound on π from a projection of Θ_I on the $\boldsymbol{\eta}$ component of $\boldsymbol{\theta}$.

4 Deductible Choices in Auto Collision Insurance

In this section and the next, we apply our theoretical findings to learn about the distributions of risk preferences and choice set size from data on households' deductible choices in auto collision insurance. In this section, we specify a random expected utility model that allows for unobserved heterogeneity in risk aversion and choice sets and describe our data.

4.1 Empirical Model

We model households' deductible choices in auto collision insurance. Each household i faces a menu of prices $\mathbf{p}_i = (p_{ic}, c \in \mathcal{D})$, where p_{ic} is the household specific premium associated with deductible c and $\mathcal{D}$ is the feasible set of deductibles, has a probability μ_i of experiencing a claim during the policy period, and has an array of observed characteristics $\mathbf{t}_i$.[11] Following the related literature (e.g., Cohen and Einav 2007; Sydnor 2010; Barseghyan et al. 2011, 2013, 2016),[12] we make two simplifying assumptions about claims and their probabilities.

Assumption 4.1 (Claims and Claim Probabilities):

(I) Households disregard the possibility of more than one claim during the policy period.

(II) Any claim exceeds the highest deductible in $\mathcal{D}$; payment of the deductible is the only cost associated with a claim; and deductible choices do not influence claim probabilities.

Assumption 4.1(I) is motivated by the fact that claim rates are small, so the likelihood of two or more claims in the same policy period is very small.[13] Assumption 4.1(II) abstracts from small claims, transaction costs, and moral hazard.

Under Assumption 4.1, household i's choice of deductible involves a choice among binary lotteries, indexed by $c \in \mathcal{D}$, of the following form: $L_i(c) \equiv (-p_{ic}, 1 - \mu_i; -p_{ic} - c, \mu_i)$. The household chooses among these lotteries based on the criterion in equation (2.1). We assume that household i's preferences conform to expected utility theory,

$$U_i(c) = (1 - \mu_i)u_i(w_i - p_{ic}) + \mu_i u_i(w_i - p_{ic} - c), \tag{4.1}$$

where w_i is the household's wealth and u_i is its Bernoulli utility function.

We impose the following shape restriction on u_i.

Assumption 4.2 (CARA): *The function u_i exhibits constant absolute risk aversion, i.e., $u_i(y) = \frac{1-\exp(-\nu_i y)}{\nu_i}$ for $\nu_i \neq 0$ and $u_i(y) = y$ for $\nu_i = 0$.*

Assuming CARA has two key virtues. First, u_i is fully characterized by the coefficient of absolute risk aversion, $\nu_i \equiv -u_i''(y)/u_i'(y)$. Second, ν_i is a constant function of wealth and hence one can estimate u_i without observing wealth. We note, however, that our approach can accommodate other shape restrictions (e.g., constant relative risk aversion) as well as non-expected utility models (e.g., the probability distortion model in Barseghyan et al. 2013).

In terms of the core model developed in Section 2, household i's observable attributes are $\mathbf{x}_i = (\mu_i, \mathbf{t}_i, \mathbf{p}_i)$, with $\mathbf{x}_{ic} = (\mu_i, \mathbf{t}_i, p_{ic})$, and its sole unobservable attribute is its coefficient of absolute risk aversion ν_i.[14] Per Assumptions 2.1 and 4.2, we posit that $\nu_i \sim P(\boldsymbol{\gamma}(\mathbf{t}_i))$, where

[11] As we explain in Section 4.2, we estimate μ_i and treat it as data.

[12] For a survey, see Barseghyan et al. (2018, Section 5.2).

[13] It also forestalls the critique that very small risks are driving our inferences about risk preferences.

[14] In terms of the notation used in Section 2, $\mathbf{s}_i = (\mu_i, \mathbf{t}_i)$, $\mathbf{z}_{ic} = p_{ic}$, and $\boldsymbol{\nu}_i = \nu_i$.

P is specified below in Assumption 4.3(I), and that, $(\mathbf{x}_{ic}, \nu_i) - a.s.$,

$$U_i(c) = \frac{(1-\mu_i)(1-\exp(\nu_i p_{ic})) + \mu_i(1-\exp(\nu_i(p_{ic}+c)))}{\nu_i}. \tag{4.2}$$

Observe that, by equation (4.2), we assume that μ_i and p_{ic} affect utility directly and we allow $\mathbf{t}_i$ to affect utility indirectly through ν_i. To capture this indirect effect, we could specify $\boldsymbol{\gamma}(\mathbf{t}_i) = f(\mathbf{t}_i; \boldsymbol{\delta})$ where the functional form of f is known up to $\boldsymbol{\delta} \in \Delta$. Instead, we account for observed heterogeneity in preferences nonparametrically by conducting the analysis separately on population subgroups based on $\mathbf{t}_i$.

Per Assumption 2.2(I), we suppose that the deductible choices and observable attributes, $\{(d_i, \mathbf{x}_i) : i \in I\}$, for a random sample of households $I \subset \mathcal{I}$, $|I| = n$, are observed, but that the households' choice sets, $\{C_i : C_i \subseteq \mathcal{D}, i \in I\}$, are unobserved. Per Assumption 2.2(II), we assume that $\Pr(|C_i| \geqslant \kappa) = 1$ for every household $i \in \mathcal{I}$, where $\kappa \geqslant 2$. At this point, however, we do not impose Assumption 3.1. Accordingly, conditional on $\mathbf{x}_i$, C_i can be arbitrarily correlated with ν_i. We impose Assumption 3.1 only in Section 5.2 when we apply Corollary 3.1 to learn about the distribution of choice set size.

We close the empirical model with two final assumptions.

Assumption 4.3 (Heterogeneity Restrictions):

(I) Conditional on $\mathbf{t}_i$, ν_i follows a Beta distribution on $[0, 0.03]$ with parameter vector $\boldsymbol{\gamma}(\mathbf{t}_i) = (\gamma_1(\mathbf{t}_i), \gamma_2(\mathbf{t}_i))$ and is independent of (μ_i, p_{ic}). To simplify notation, we suppress below the dependence of $\boldsymbol{\gamma}$ on $\mathbf{t}_i$.

(II) The minimum choice set size is $\kappa = 3$.

Assumption 4.3(I) specifies that P is the Beta distribution with support $\mathcal{V} = [0, 0.03]$. The main attraction of the Beta distribution is its flexibility (e.g., Ghosal 2001). Its bounded support is a plus given our setting. A lower bound of zero rules out risk loving preferences and seems appropriate for insurance markets that exist primarily because of risk aversion. Imposing an upper bound enables us to rule out absurd levels of risk aversion, and the choice of 0.03 is conservative both as a theoretical matter and in light of prior empirical estimates in similar settings (e.g., Cohen and Einav 2007; Sydnor 2010; Barseghyan et al. 2011, 2013, 2016). Assumption 4.3(II) posits that the size of every household's choice set is either full, full-1, or full-2. In our setting $|\mathcal{D}| = 5$. We set $\kappa = 3$ for reasons we explain in Section 4.2.

Remark 4.1: We also consider a mixed logit specification $U_i(c) = \omega(\mathbf{x}_{ic}, \nu_i) + \epsilon_{ic}$, where $\omega(\mathbf{x}_{ic}, \nu_i)$ is the certainty equivalent of the right hand side of equation (4.2), ν_i is distributed per Assumption 4.3(I), and ϵ_{ic} is an i.i.d. disturbance that follows a Type 1 Extreme Value distribution and is independent of $(\mathbf{x}_{ic}, \nu_i)$; see Section 5.1.1.

4.2 Data Description

We obtained the data from a large U.S. property and casualty insurance company. The data contain annual information on more than 100,000 households who first purchased auto policies from the company during the ten year period from 1998 to 2007. We focus on households' deductible choices in auto collision coverage. This coverage pays for damage to the insured vehicle, in excess of the deductible, caused by a collision with another vehicle or object, without regard to fault. The feasible set of auto collision deductibles is $\mathcal{D} = \{\$100, \$200, \$250, \$500, \$1000\}$ and thus $|\mathcal{D}| = 5$.

To construct our analysis sample, we initially include every household who first purchased auto collision coverage from the company between 1998 and 2007, retaining, at the time of first purchase, its deductible choice d_i, its pricing menu $\mathbf{p}_i$, its claim probability μ_i, and an array $\mathbf{t}_i$ of three demographic characteristics: gender, age, and insurance score of the principal driver.[15] This yields an initial sample of 112,011 households. We then exclude households whose deductible choices cannot be rationalized by the model specified in Section 4.1 for any pair (ν_i, C_i) such that $\nu_i \in [0, 0.03]$ and $|C_i| \in \{3, 4, 5\}$. Importantly, our rationalizability check does *not* rely on the assumption that P is the Beta distribution. This excludes 0.1 percent of the initial sample, yielding a final sample of 111,890 households.[16]

Several comments are in order. First, we retain households' deductible choices at the time of first purchase to increase confidence that we are working with active choices. One might worry that households renew their policies without actively reassessing their deductibles.

Second, we require $\nu_i \in [0, 0.03]$ for the reasons stated in Section 4.1. However, the composition of our sample is robust to the upper bound of the support. If we decrease the upper bound to 0.02 the sample decreases by one household to 111,889 households. If we increase the upper bound to 0.04 the sample remains the same at 111,890 households.[17]

Third, we require $|C_i| \in \{3, 4, 5\}$—i.e., we assume $\kappa = 3$—to keep the model as close as possible to the standard approach that assumes full size choice sets. As we explain in Section 5.2, $\kappa = 3$ is the highest value that is consistent with the data.

[15] Insurance score is a credit based risk score.

[16] The data in this paper are not the same as the data in Barseghyan et al. (2013) and Barseghyan et al. (2016), though both data sets have the same source. In this paper, the data comprise 112,011 households who first purchased auto collision coverage between 1998 and 2007. In Barseghyan et al. (2013) and Barseghyan et al. (2016), the data comprise 4,170 households who first purchased auto collision coverage, auto comprehensive coverage, and home all perils coverage in the same year, in either 2005 or 2006.

[17] Moreover, our results are robust to increasing the upper bound from 0.03 to 0.04, as indicated by results available from the authors upon request.

Fourth, the company generates each household's pricing menu, $\mathbf{p}_i = (p_{ic}, c \in \mathcal{D})$, according to the following pricing rule: $p_{ic} = g(c)\bar{p}_i + \zeta$, where $\bar{p}_i$ is the household's base price, g is a decreasing positive function, and $\zeta > 0$. We observe g, ζ, and the premium paid by each household given its chosen deductible. We thus can recover each household's base price. Given the company's pricing rule, the base price is a sufficient statistic for $\mathbf{p}_i$. Moreover, any $p_{ic} \in \mathbf{p}_i$ can be treated as the base price. We treat the premium associated with the \$1000 deductible as the base price—i.e., $\bar{p} = p_{1000}$—and round it to the nearest five dollars. We use the rounded base prices and resulting pricing menus throughout our analysis.[18]

Fifth, 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 deductible choices do not influence claim rates (Assumption 4.1(II)). We perform a Poisson panel regression with random effects and use the results to calculate a fitted claim rate $\widehat{\lambda}_i$ for each household.[19] 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.[20] We treat μ_i as data.

Table 4.1 presents descriptive statistics for the analysis sample. Panel A summarizes the households' deductible choices, pricing menus, claim probabilities, and demographic characteristics. Panel B reports the sample distribution of deductible choices for the full sample and for subsamples based on gender, age, and insurance score.[21] In Table 4.1 and throughout the paper, young/old and low/high insurance scores are defined as bottom/top third based on the age and insurance score, respectively, of the principal driver.

[18]This includes our rationalizability check, though the final sample would be virtually identical if we used exact prices. Our use of rounded prices reduces the computational burden of recovering Θ_I and is supported by evidence that "people show a marked tendency to produce 0- and 5-ending numbers" in numerical cognition tasks, including price cognition (Schindler and Kirby 1997, p. 193). See also Schindler and Wiman (1989), Vanhuele and Drèze (2000), and Liang and Kanetkar (2006).

[19]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. As we explain in Section C.1 of the Supplemental Material, we calculate $\widehat{\lambda}_i$ conditional on the household's observables at the time of first purchase and its subsequent claims experience.

[20]Our use of rounded claim probabilities reduces the computational burden of recovering Θ_I and is supported by evidence that people report rounded probabilities (Manski and Molinari 2010).

[21]In addition, Table C.1 in the Supplemental Material reports the sample distribution of deductible choices by octiles of base price $\bar{p}$ and claim probability μ.

Table 4.1: Descriptive Statistics

Panel A. Summary Statistics	Mean	Std. dev.	5th pctl.	Median	95th pctl.
Deductible choice (dollars)	439	178	200	500	500
Pricing menus:					
p_{500}	217	137	77	181	480
$p_{250} - p_{500}$	65	42	22	54	146
$p_{500} - p_{1000}$	49	32	17	41	110
Claim probability (annual)	0.088	0.030	0.045	0.085	0.140
Demographic characteristics:					
Female	0.468	0.499	0	0	1
Age (years)	48.1	16.6	24.5	45.9	76.7
Insurance score	731	114	555	725	934

Panel B. Deductible Choices		Percent choosing deductible				
	Obs.	\$100	\$200	\$250	\$500	\$1000
All households	111,890	1.1	15.2	13.7	65.4	4.6
Male	59,476	1.0	14.9	12.9	65.9	5.4
Female	52,414	1.1	15.5	14.7	64.8	3.8
Young	36,932	0.1	6.9	10.7	77.1	5.2
Old	38,046	2.5	26.2	16.7	51.0	3.6
Low Insurance Score	37,087	0.4	10.1	12.7	72.2	4.6
High Insurance Score	38,371	1.8	20.9	14.6	58.1	4.6

Notes: Analysis sample (111,890 households). Pricing statistics are annual amounts in nominal dollars. Demographic statistics are for the principal driver.

5 Empirical Findings

Our empirical application is motivated in part by the fact that, although we observe the feasible set of deductibles, we do not observe which deductibles enter a household's choice set. There are many plausible sources of unobserved heterogeneity in choice sets. It may be due to missing data—e.g., different sales agents may quote different subsets of deductibles to different households—or to unobserved constraints—e.g., some households may disregard low deductibles due to budget constraints or high deductibles due to liquidity constraints.

Our application is also motivated by a persistent finding in prior empirical studies of risk preferences which assume full size choice sets. These studies tend to find that average risk aversion is quite high—arguably implausibly high. Two recent examples that utilize similar data are Cohen and Einav (2007) and Barseghyan et al. (2013). We suspect that the assumption of full size choice sets may be driving this finding and that allowing for unobserved heterogeneity in choice sets may yield more credible estimates of risk preferences.

In what follows we first apply Theorem 3.1, which does *not* assume independence between preferences and choice sets, to learn about the distribution of risk aversion (Section 5.1). We then apply Corollary 3.1, which assumes that choice set size is independent of preferences (Assumption 3.1), to learn about the distribution of choice set size (Section 5.2).

In the text we present results for the population (all households). As indicated in Section 4.1, we also conduct our analysis separately for population subgroups based on observed characteristics $\mathbf{t}_i$. The subgroup results are reported in Section C.3 of the Supplemental Material. For the population and each subgroup, the conditional moment inequalities in equations (3.5) and (3.7) need to hold $(\mu_i, \bar{p}_i) - a.s.$ We therefore compute, for the population and each subgroup, a confidence set that asymptotically uniformly covers the vector $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ with probability 95 percent,[22] using the method proposed by Andrews and Shi (2013) [hereafter, AS].[23] In addition, we report 95 percent confidence intervals for percentiles of ν_i based on projections of the AS confidence set, and we apply the bootstrap-based calibrated projection method proposed by Kaido et al. (2019) [hereafter, KMS] to obtain asymptotically uniformly valid 95 percent confidence intervals for $\mathrm{E}(\nu_i)$, π_3, π_4, and π_5.[24] We review the AS and KMS methods in Section B of the Supplemental Material.[25]

Following AS, we aggregate the inequalities in equations (3.5) and (3.7), as the case may be, by discretizing the support of $(\mu_i, \bar{p}_i)$ into 65 "hypercubes."[26] Where applicable, we leverage the strategies referenced in Section 3.1.1 to reduce our computational burden. We also mention that there are values of $\boldsymbol{\theta} \in \Theta_I$ for which the sample analogs of the moment inequalities in equations (3.5) and (3.7) are satisfied. This implies that we fail to reject the hypothesis that our empirical model is correctly specified.[27]

To provide context for our risk aversion estimates in Section 5.1, we also report estimates obtained under two point identified expected utility models. They are:

- Uniform Random (UR): Utility is given by equation (4.2). Choice sets are drawn uniformly at random from $\mathcal{D}$, conditional on $|C_i| = q$ for $q \geqslant \kappa$ and independent of ν_i. Specifically, $\Pr(C_i = G||G| = q) = \binom{|\mathcal{D}|}{q}^{-1}$ for all $G \subseteq \mathcal{D}$, $|G| = q$, $q \geqslant \kappa$; and $C_i \perp \nu_i$.

[22]For $\nu_i \sim Beta(\gamma_1, \gamma_2)$, a unique pair $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ corresponds to each pair (γ_1, γ_2).

[23]The AS confidence set asymptotically exploits all the information in the conditional moments, in the sense that as the sample size grows to infinity the number of inequalities used for inference increases and the confidence set shrinks to the (group specific) sharp identification region.

[24]Although they do not asymptotically exploit all the information in the conditional moments because they are based on a fixed number of inequalities, the KMS confidence intervals (implemented on the same sample with the same inequalities and tuning parameters) are shorter than those obtained by projecting the AS confidence set.

[25]Both the AS and KMS methods entail the selection of tuning parameters. We find that our results are robust to the choice of tuning parameters, as indicated by results available from the authors upon request.

[26]Following AS, μ_i and $\bar{p}_i$ are normalized by their covariance matrix to ensure more uniform hypercubes. Each hypercube contains between 660 and 2,901 households, except for one that contains all households.

[27]For methods to test for misspecification in moment inequality models, see Bugni et al. (2015).

- Alternative Specific Random (ASR): Utility is given by equation (4.2). Alternatives in $\mathcal{D}$ enter choice sets with alternative specific probabilities, independent of one another and ν_i, conditional on $|C_i| \geqslant \kappa$ (cf. Manski 1977; Manzini and Mariotti 2014). Specifically, $\Pr(C_i = G||G| \geqslant \kappa) = \Pr(C_i = G)/(1 - \sum_{G \subseteq \mathcal{D}:|G|<\kappa} \Pr(C_i = G))$ for all $G \subseteq \mathcal{D}$, where $\Pr(C_i = G) = \prod_{c \in G} \varphi(c) \prod_{c \in \mathcal{D} \setminus G}(1 - \varphi(c))$ and $\varphi(c) = \Pr(c \in C_i)$; and $C_i \perp \nu_i$.

UR and ASR are "reduced form" models that can capture a wide range of choice set formation processes. For example, UR is consistent with a simultaneous search process with a uniform prior (cf. Stigler 1961),[28] and ASR may describe an advertising process in which alternatives are marketed with different intensities in independent, non-targeted campaigns. With dependence between $\varphi(c)$ and ν_i, ASR can capture an even wider range of choice set formation processes, including, for instance, a sequential search process with free recall (e.g., Weitzman 1979) or an advertising process with correlated, targeted campaigns.

5.1 Risk Preferences

Panel (a) of Figure 5.1 depicts the AS 95 percent confidence set for $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ for all households.[29] In addition, Table 5.1 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 by each bound. For the percentiles, we report only the implied risk premia. Focusing on the lower bounds, the main takeaway is that the households' deductible choices can be explained by a distribution of absolute risk aversion that has a low mean, on the order of 10^{-3}, and low variance, on the order of 10^{-6}. Strikingly, the lower bound on the 25th percentile of ν_i corresponds to a risk premium of less than half a cent, implying that the data are consistent with at least a quarter of households being effectively risk neutral.

To provide context for these results, Table 5.1 also reports: (i) 95 percent confidence intervals for the mean, 25th percentile, and 75th percentile of ν_i obtained under UR and ASR; and (ii) point estimates for the mean of ν_i reported by Cohen and Einav (2007) and Barseghyan et al. (2013) for their CARA models. Cohen and Einav (2007) estimate the distribution of absolute risk aversion in a parametric expected utility model using data on deductible choices in Israeli auto insurance. Barseghyan et al. (2013) estimate the distribu-

[28]With a uniform prior, the simultaneous search problem reduces to choosing the optimal number of alternatives to search and, given this number, randomly choosing the alternatives to be searched.

[29]In Figure C.2 in the Supplemental Material, we also report a 95 percent confidence set for an outer region of admissible probability density functions of ν_i.

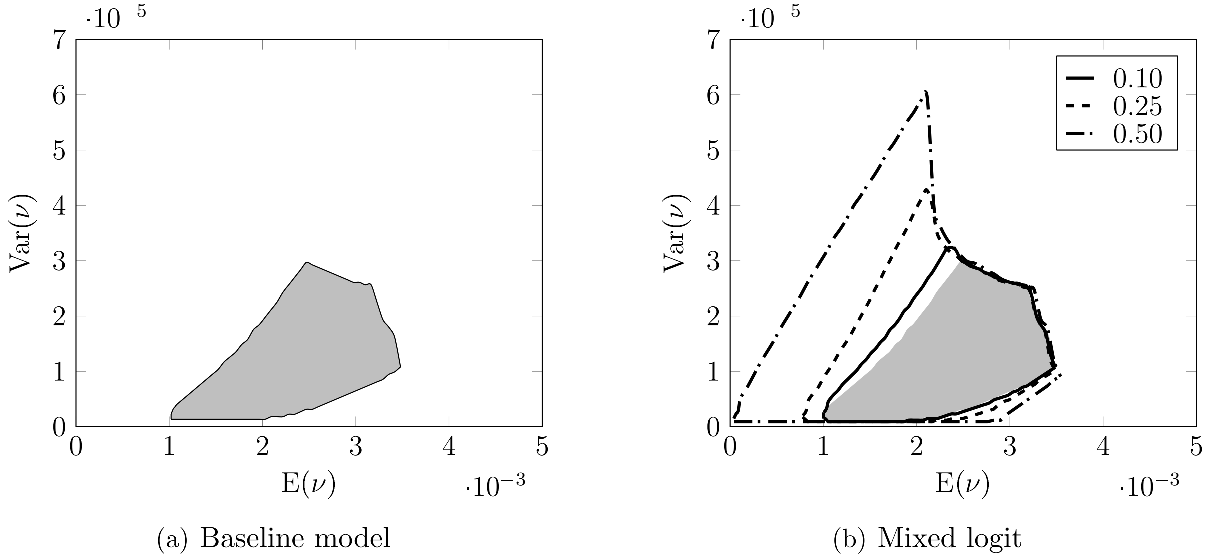


Figure 5.1: AS 95 percent confidence sets for $(\mathrm{E}(\nu), \mathrm{Var}(\nu))$.

tions of absolute risk aversion and probability distortions in a parametric rank-dependent expected utility model using data on deductible choices in U.S. auto and home insurance.

The main takeaway is that the baseline lower bounds are substantially smaller than the lower bounds obtained under UR and ASR and the point estimate reported by Cohen and Einav (2007). This suggests that if one properly allows for heterogeneity in choice sets, the data can be explained by expected utility theory with substantially lower levels of risk aversion than many familiar models—including some that allow for choice set heterogeneity but perhaps misspecify the choice set formation process—would imply. A second takeaway comes from results in Barseghyan et al. (2013). Their point estimate for the mean of ν_i is only slightly larger than the baseline lower bound ($68 versus $62 in terms of implied risk premium). However, because they allow for probability distortions, ν_i does not fully capture a household's level of risk aversion in their model. Taking into account their point estimate for probability distortions, the implied risk premium is $91, suggesting that heterogeneous choice sets have very different behavioral implications than probability distortions.

5.1.1 Mixed Logit with Unobserved Heterogeneity in Choice Sets

We also compute the AS 95 percent confidence set for $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ for a mixed logit specification $U_i(c) = \omega(\mathbf{x}_{ic}, \nu_i) + \epsilon_{ic}$, where $\omega(\mathbf{x}_{ic}, \nu_i)$ is the certainty equivalent of the right hand side of equation (4.2), ν_i is distributed per Assumption 4.3(I), and ϵ_{ic} is an i.i.d. disturbance that follows a Type 1 Extreme Value distribution with scale parameter σ and is independent of $(\mathbf{x}_{ic}, \nu_i)$. We define utility in terms of its certainty equivalent so that ϵ_{ic} is measured in

Table 5.1: Distribution of Absolute Risk Aversion

	Mean		Implied risk premium					
			Mean		25th pctl.		75th pctl.	
	LB	UB	LB	UB	LB	UB	LB	UB
Baseline model	0.00105	0.00347	\$ 62	\$307	\$ 0	\$ 78	\$ 79	\$454
UR	0.00167	0.00170	\$115	\$117	\$ 86	\$ 88	\$142	\$145
ASR	0.00260	0.00264	\$211	\$216	\$ 40	\$ 43	\$333	\$340
Cohen and Einav (2007)	0.00310		\$267		Not reported		Not reported	
Barseghyan et al. (2013)	0.00113		\$ 68		Not reported		Not reported	

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

dollars (which allows for a clear economic interpretation).[30] Panel (b) of Figure 5.1 depicts the confidence set for three values of σ chosen so that the standard deviation of ϵ_{ic} is equal to 10 percent, 25 percent, and 50 percent of the average price difference among adjacent deductibles in $\mathcal{D}$. (At zero percent, of course, the mixed logit specification reduces to the baseline model.) As the "noise factor" increases, the confidence set expands mainly to the "northwest," admitting higher values of $\mathrm{Var}(\nu_i)$ and lower values of $\mathrm{E}(\nu_i)$. Focusing on the latter, the projection of the confidence set on $\mathrm{E}(\nu_i)$ is essentially unchanged at a noise factor of 10 percent. At 25 percent the lower bound is smaller but still informative. By 50 percent, however, the confidence set effectively admits $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i)) = (0, 0)$ (i.e., all households are risk neutral) and overall is quite large. The bottom line is that the confidence set remains informative at reasonable levels of noise. Not surprisingly, however, as the magnitude of the noise approaches that of the variation in observable covariates, the data loses much of its informational content about preferences.

5.2 Choice Set Size

Table 5.2 reports KMS 95 percent confidence intervals for π_5, π_4, and π_3. 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

[30]Moreover, if we did not use certainty equivalents, the mixed logit specification would be subject to the nonmonotonicity critique of Apesteguia and Ballester (2018, Corollary 1), who show that random expected utility models with CARA or CRRA utility and additive i.i.d. disturbances violate a basic monotonicity property: given any choice set, as risk aversion increases the choice probabilities of the riskier alternatives decrease at first but eventually increase (because differences in expected utilities converge to zero as risk aversion increases, allowing differences in disturbances to determine choices). This is not the case for our baseline model, which is a random parameter model and thus is immune to their critique even without using certainty equivalents (Apesteguia and Ballester 2018, Proposition 5). Although monotonocity problems can still arise with certainty equivalents, our lotteries do not run afoul of their nonmonotonicity result for certainty equivalents (Apesteguia and Ballester 2018, Corollary 2).

Table 5.2: Distribution of Choice Set Size

	π_5 (full)		π_4 (full-1)		π_3 (full-2)	
	LB	UB	LB	UB	LB	UB
All households	0.00	0.24	0.00	0.89	0.11	1.00

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

with full-1 choice sets. By implication, one minus the former is the minimum fraction of households who require full-1 or full-2 choice sets to rationalize their deductible choices, while one minus the latter (which equals the lower bound on π_3) is the minimum fraction of households who require full-2 choice sets.[31]

The main result is that a large majority of households require limited choice sets (full-1 or full-2) to explain their deductible choices. Specifically, we find that at least 76 percent of households require limited choice sets, including at least 11 percent who require full-2 choice sets. In the remainder of this section we discuss two drivers of this key result: suboptimal choices and violations of the law of demand.[32]

5.2.1 Suboptimal Choices

The first driver is the existence and frequency of suboptimal choices. In total, 16.7 percent of households in our sample choose a deductible that is suboptimal (i.e., not first best in $\mathcal{D}$) under our empirical model at all $\nu \in [0, 0.03]$. The vast majority of these households choose \$200, which is a suboptimal alternative under the model for virtually every household in our sample.[33] In particular, \$200 is dominated by \$100 or \$250, depending on μ. Suboptimal alternatives, sometimes called dominated alternatives, are not uncommon in discrete choice settings, including insurance settings (see, e.g., Handel 2013; Bhargava et al. 2017).

To see why \$200 is a suboptimal alternative under the model, consider a risk neutral household with claim probability μ. The household prefers \$200 to \$100 if and only if $\mu < \frac{p_{100}-p_{200}}{200-100}$, and prefers \$200 to \$250 if and only if $\mu > \frac{p_{200}-p_{250}}{250-200}$. In our data $p_{100} - p_{200} = p_{200} - p_{250}$ for all households. For the risk neutral household, therefore, at most one of the foregoing inequalities holds and thus \$200 is dominated by \$100 or \$250, depending on the value of μ. A similar logic applies for risk averse households with reasonable levels of risk aversion—under our model or any other model in which lotteries are evaluated by

[31]By construction, because $\kappa = 3$, the lower bounds on π_5 and π_4 are zero, the lower bound on π_3 is one minus the upper bound on π_4, and the upper bound on π_3 is one.

[32]In other applications, different or additional data features may reveal the presence of heterogeneous choice sets. One example is zero shares for alternatives that are not suboptimal.

[33]The remainder of these households choose \$1000 or \$500 when \$250 is optimal.

expectations over functions of final wealth (see Barseghyan et al. 2016)—and indeed for virtually every household in our sample \$200 is suboptimal at all $\nu \in [0, 0.03]$.[34]

Yet 15.2 percent of households in our sample choose \$200. At the same time, only 1.1 percent choose \$100 and 13.7 percent choose \$250. Hence, the combined demand for \$100 and \$250 is less than the demand for \$200. This pattern is even more pronounced within certain subgroups, including households with old principal drivers and households with high insurance scores; see Table 4.1.

Heterogeneous choice sets can readily explain these choice patterns. In our model all that is required to rationalize a household's choice of \$200 is the absence of \$100 or \$250, as the case may be, from the household's choice set. Moreover, all that is required to explain $\Pr(d = 100|\mathbf{x}) + \Pr(d = 250|\mathbf{x}) > \Pr(d = 200|\mathbf{x})$ is a choice set distribution in which the frequencies of \$100 and \$250 are sufficiently less than the frequency of \$200.

With full size choice sets, however, our model cannot explain these choice patterns. 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.

PROPERTY 5.1 (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* $W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) \geqslant W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})$, $\boldsymbol{\nu} - a.s.$.

Indeed, *any* model that satisfies an analogous property is incapable of explaining the relative frequency of \$200 in the distribution of observed deductible choices.[35] This includes, inter alia, the conditional and mixed logit models (McFadden 1974; McFadden and Train 2000), the semiparametric random utility model of Manski (1975), and the multinomial probit model (e.g., Hausman and Wise 1978).[36] At the same time, not all choice set formation processes can explain these choice patterns. For instance, UR cannot but ASR can.

CLAIM 5.1: *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 5.1 implies $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) > \Pr(d = c|\mathbf{x})$, $\mathbf{x} - a.s.$

[34] Evaluating equation (4.2) for all 111,890 households over a fine grid of ν, we find that the \$200 deductible is optimal in 0.001 percent of cases, all of which entail $\nu \geqslant 0.0115$.

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

[36] This also includes the probability distortion model in Barseghyan et al. (2016), which explains why they find that 13.0 percent of the households in their data cannot be rationalized by their model.

(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.*

The proof of Claim 5.1(I) is set forth in Section C.5 of the Supplemental Material.[37] We emphasize that Claim 5.1 does not rely on Assumption 3.1 or the assumptions of the empirical model set forth 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.

5.2.2 Law of Demand

Violations of the law of demand are also driving our main result on choice sets. With full size choice sets, households' demand for high deductibles should increase as base price increases and should decrease as claim risk increases. If follows that, with full size choice sets, we should observe for all $K \in \overleftarrow{K} \equiv \big\{\{\$1000\}, \{\$1000, 500\}, \{\$1000, \$500, \$250\}\big\}$,

$$\Pr(d \in K|\mu, \bar{p}) > \Pr(d \in K|\mu', \bar{p}') \text{ if } \mu < \mu' \text{ and } \bar{p} > \bar{p}'. \tag{5.1}$$

In our data, however, we observe multiple violations. In particular, when we compare all pairs of hypercubes, where one hypercube has a lower average μ and a higher average $\bar{p}$ than the other, over all subsets $K \in \overleftarrow{K}$, we find 61 violations (3 percent) of equation (5.1).[38]

The requirement in equation (5.1) holds generically for models in which $\frac{\partial[U(c)-U(c')]}{\partial \bar{p}} > 0$ and $\frac{\partial[U(c)-U(c')]}{\partial \mu} < 0$ for all $c, c' \in \mathcal{D}$, $c > c'$. Given the assumptions of our empirical model, the law of demand implies a second, stronger requirement (cf. Barseghyan et al. 2020). Observe that for any $\mathbf{x} = (\mu, \bar{p})$ and any subset $K \subset \mathcal{D}$ of adjacent deductibles, there exists an interval $\mathcal{S}_K(\mathbf{x}) \subseteq \mathcal{V}$ such that $d^*(\mathcal{D}; \mathbf{x}, \nu) \in K$ if $\nu \in \mathcal{S}_K(\mathbf{x})$ and $d^*(\mathcal{D}; \mathbf{x}, \nu) \in \mathcal{D}\backslash K$ if $\nu \in \mathcal{V}\backslash\mathcal{S}_K(\mathbf{x})$, where $d^*(\mathcal{D}; \mathbf{x}, \nu)$ denotes the model implied optimal choice when the choice set has full size. It follows that, with full size choice sets,

$$\Pr(d \in K|\mathbf{x}) \leqslant \Pr(d \in K'|\mathbf{x}') \text{ if } \mathcal{S}_K(\mathbf{x}) \subset \mathcal{S}_{K'}(\mathbf{x}') \tag{5.2}$$

for any subsets $K, K' \subset \mathcal{D}$ of adjacent deductibles and any $\mathbf{x} \neq \mathbf{x}'$. In our data, however, we observe numerous violations of equation (5.2). In particular, when we compare all pairs of hypercubes, where $\mathbf{x}$ denotes the average $(\mu, \bar{p})$ in one hypercube and $\mathbf{x}'$ denotes the average $(\mu', \bar{p}')$ in the other, over all subsets $K, K' \subset \mathcal{D}$ of adjacent deductibles where each subset

[37] An analogous claim holds in the case of a model with additively separable noise 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 $\omega(\mathbf{x}_b, \boldsymbol{\upsilon}; \boldsymbol{\delta}) > \omega(\mathbf{x}_c, \boldsymbol{\upsilon}; \boldsymbol{\delta})$.

[38] We do not count violations where K contains a suboptimal alternative under the model given the average $(\mu, \bar{p})$ in either hypercube.

contains either one, two, or three deductibles, we find 44,847 instances (15 percent) in which $\mathcal{S}_K(\mathbf{x}) \subset \mathcal{S}_{K'}(\mathbf{x}')$ but $\Pr(d \in K|\mathbf{x}) > \Pr(d \in K'|\mathbf{x}')$.[39]

We conclude by highlighting how equation (5.2) relates to the characterization of Θ_I in Corollary 3.1. Consider whether any parameter vector with $\pi_{|\mathcal{D}|} = 1$ belongs to Θ_I. At that value $D^*_\kappa(\mathbf{x}, \nu) = \{d^*(\mathcal{D}; \mathbf{x}, \nu)\}$, a singleton, and hence the inequality in equation (3.7), evaluated at any subset $K \subseteq \mathcal{D}$ of adjacent deductibles and its complement $\mathcal{D} \backslash K$, implies

$$\Pr(d \in K|\mathbf{x}) = \sum_{c \in K} \int \mathbf{1}(d^*(\mathcal{D}; \mathbf{x}, \tau) = c) dP(\tau; \boldsymbol{\gamma}) = \int_{\mathcal{S}_K(\mathbf{x})} dP(\tau; \boldsymbol{\gamma}),$$

which in turn implies equation (5.2). Thus, a violation of equation (5.2) implies that no parameter vector with $\pi_{|\mathcal{D}|} = 1$ belongs to Θ_I. A similar logic applies to the choice probabilities of suboptimal alternatives. In general, our method—through the inequalities in equation (3.7)—takes into account all restrictions implied by the data and the model, while accounting for finite sample uncertainty.

6 Discussion

Discrete choice analysis in the tradition of McFadden (1974) contemplates heterogeneity in agents' choice sets. It however assumes that choice sets are observed by the econometrician (McFadden 1974, p. 107). In practice choice sets are often unobserved. Manski (1977, p. 239) suggests the following characterization of the outcome probability of the discrete choice process—i.e., the probability that an agent with observable attributes $\mathbf{x}_i$ and choice set G chooses alternative c— when agents' choice set are unobserved:

$$\Pr(d_i = c|\mathbf{x}_i) = \sum_{G \subseteq \mathcal{D}} \Pr(c \in^* G|\mathbf{x}_i) \Pr(C_i = G|\mathbf{x}_i, c \in G), \tag{6.1}$$

where $\in^*$ denotes "is chosen from" and $\Pr(C_i = G|\mathbf{x}_i, c \in G)$ is the probability that G is drawn from the feasible set $\mathcal{D}$ given that c is in the realized choice set.

The two-stage characterization in equation (6.1) forms the basis of numerous models of discrete choice with unobserved heterogeneity in choice sets, including ours (as one can readily see from equation (3.2) and where $\Pr(C_i = G|\mathbf{x}_i, c \in G)$ can depend on preferences). It also makes plain the nature of the identification problem when choice sets are unobserved (which we elaborate in Section 3.1). In order to point identify the model of preferences, which is represented by $\in^*$ in equation (6.1), the econometrician has to make assumptions—either explicitly or implicitly, sometimes arbitrary and often unverifiable—about the choice

[39] Again, we do not count violations where K or K' contains a suboptimal alternative under the model given $\mathbf{x}$ or $\mathbf{x}'$, respectively.

set formation process, including with respect to the dependence or lack thereof between preferences and choices sets (conditional on observables) (cf. Ben-Akiva 1973, pp. 84-85).

In what follows we provide an overview of the assumptions made in the econometrics and applied literatures on discrete choice analysis to grapple with the identification problem created by unobserved heterogeneity in choice sets.[40] We describe four prominent approaches and provide examples of recent papers that take each approach. We do not provide a comprehensive review of the literature, which is vast and spans a diverse array of fields in economics. However, our overview of the landscape enables us to situate our approach within the literature and provides context for our contributions, which we recap at the end.

* * *

The most common approach in the discrete choice literature to the identification problem created by unobserved choice sets is to assume that agents' choice sets all comprise the feasible set or a known subset of the feasible set.[41] This is the approach taken by, for example, Berry et al. (1995) in estimating demand curves from aggregate data on U.S. auto sales; Cohen and Einav (2007) in estimating risk preferences from individual-level data on deductible choices in Israeli auto insurance; and Chiappori et al. (2019) in estimating risk preferences from aggregate betting data on U.S. horse races. We also take this approach in prior work on estimating risk preferences from individual-level data on deductible choices in U.S. auto and home insurance (Barseghyan et al. 2011, 2013, 2016).

Papers that allow for heterogeneity in choice sets take three basic approaches to identification. The first is to rely on auxiliary information about the composition or distribution of agents' choice sets. For instance, Draganska and Klapper (2011), who study ground coffee sales, use survey data on brand awareness;[42] De los Santos et al. (2012), who study online book purchases, use survey data on web browsing;[43] Conlon and Mortimer (2013), who study vending machine sales, utilize periodic inventory snapshots; and Honka and Chintagunta (2017), who study auto insurance purchases, use survey data on price quotes.[44]

---

[40]Many important papers in the theory literature—including papers on revealed preference analysis under limited attention, limited consideration, and other forms of bounded rationality that manifest in unobserved heterogeneity in choice sets—also grapple with the identification problem (e.g., Masatlioglu et al. 2012; Manzini and Mariotti 2014; Caplin and Dean 2015; Lleras et al. 2017; Cattaneo et al. 2020). However, these papers generally assume rich datasets—e.g., observed choices from every possible subset of the feasible set—that often are not available in applied work, especially outside of the laboratory. A notable exception is Dardanoni et al. (2020), which assumes that only a single cross-section of aggregate choice shares is observed.

[41]Cf. Swait (2001, p. 643): "The most common strategy of choice set specification makes all choice sets equal to the master set...."; Honka et al. (2017, p. 615): "[M]ost demand side models maintain the full information assumption that consumers are aware of and consider all available alternatives."

[42]In a similar vein, Honka et al. (2017), who study bank account openings, use survey data on brand awareness and search activity.

[43]Similarly, Kim et al. (2010), who study online camcorder sales, use market data on web searches.

[44]For earlier papers, see, e.g., Roberts and Lattin (1991) and Ben-Akiva and Boccara (1995).

The second approach is to rely on two-way exclusion restrictions—i.e., assume that certain variables impact choice sets but not preferences and vice versa. For example, Goeree (2008) assumes that media advertising affects the set of computers of which a consumer is aware (and hence her choice set) but not her preferences over computers, while computer attributes affect her preferences but not her choice set;[45] Gaynor et al. (2016) assume that waiting times and mortality rates directly impact a patient's preferences over hospitals but not her referring physician's preferences (which determine her choice set), while distance to hospital and hospital fixed effects directly impact her referring physician's preferences (and hence her choice set) but not her preferences; and Hortaçsu et al. (2017) assume that a retail electricity customer's decision to consider alternatives to her retailer is a function of her last period retailer (e.g., a bad customer service experience) but not her next period retailer, while her choice of retailer is a function of her next period retailer but not her last period retailer.[46]

The last approach is to rely primarily on restrictions to the choice set formation process. Five recent papers that exemplify this approach are Abaluck and Adams (2020), Barseghyan et al. (2020), Crawford et al. (2020), Lu (2019), and Cattaneo et al. (2020).[47]

Abaluck and Adams (2020) consider two models of choice set formation: a variant of the ASR model described above and a "default specific" model in which each agent's choice set comprises either a single, default alternative or the entire feasible set. They show that the restrictions imposed on choice probabilities by these models are sufficient for point identification of preferences and choice set probabilities due to induced asymmetries in cross-attribute responses ('Slutsky asymmetries'), assuming that choice sets and preferences are independent conditional on observables and that every alternative has a continuous attribute with large support that is additively separable in utility and shifts choice set probabilities.

Barseghyan et al. (2020) study point identification of discrete choice models with unobserved heterogeneity in preferences and choice sets. They establish conditions for point identification of the preference distribution under generic choice set formation processes. They also illustrate the tradeoff between the common exclusion restrictions and the restrictions on choice set formation required for semi-nonparametric point identification.

Crawford et al. (2020) show that with panel data (or group-homogeneous cross-section data) and preferences in the logit family, point identification of preferences is possible, with-

[45]Similarly, van Nierop et al. (2010) assume that in-store marketing impacts which brands of laundry detergent and yogurt a shopper considers (and hence her choice set) but not her preferences over brands, while brand attributes impact her preferences but not her choice set.

[46]Heiss et al. (2019) similarly assume that a Medicare Part D insured's decision to consider alternatives to her existing drug plan is triggered by past changes in her plan's attributes (e.g., a price increase), while her plan choice is determined by current attributes of available plans. See also Ho et al. (2017).

[47]Dardanoni et al. (2020) also take this approach. However, they rule out unobserved preference heterogeneity and focus on point identification of the choice set formation model.

out any exclusion restrictions, under the assumption that choice sets and preferences are independent conditional on observables and with restrictions on how choice sets evolve over time. These restrictions enable the construction of proper subsets of agents' true choice sets ('sufficient sets') that can be utilized to estimate the preference model.

Lu (2019) provides conditions for both partial and point identification of a random coefficient logit model. He assumes that each agent's unobserved choice set is bounded by two observed sets, her largest possible choice set (e.g., the feasible set) and her smallest possible choice set (containing a default alternative and at least one other alternative). He shows that availability of these data, together with the assumption that agents' choices obey Sen's property α, yields moment inequalities on the choice probabilities, which he uses to obtain outer regions on the model's preference parameters.

Cattaneo et al. (2020) propose a random attention model in which agents' preferences are homogeneous (and thus independent of choice sets) and the probability of a particular choice set does not decrease when the number of possible choice sets decreases. Within this framework, they provide revealed preference theory and testable implications for observable choice probabilities, as well as partial identification results for preference orderings.

The approach that we propose and apply in this paper falls into this last category. However, it relies on fewer and weaker restrictions on the choice set formation process than any other paper in that category. Our core model imposes—and hence our main identification result requires—only one mild assumption on the choice set formation process, namely that choice sets have a known minimum size greater than one. Importantly, our core model does not assume that choice sets are independent of preferences conditional on observables (Abaluck and Adams 2020; Crawford et al. 2020; Cattaneo et al. 2020). Nor do we impose other restrictions on how agents' choice sets are formed (Abaluck and Adams 2020; Barseghyan et al. 2020) or evolve over time (Crawford et al. 2020), rely on exclusion restrictions or large support assumptions (Abaluck and Adams 2020; Barseghyan et al. 2020), require that the econometrician knows the composition of the smallest possible choice set for each agent (Abaluck and Adams 2020; Lu 2019), or assume that choice sets satisfy a monotonicity or other regularity condition (Lu 2019; Cattaneo et al. 2020).

Due to the parsimony of our approach we obtain partial and not point identification of the underlying model of preferences. Nevertheless, we demonstrate that much can be learned about the distribution of preferences under our approach. Moreover, what is learned has more credibility because we avoid making a host of arbitrary or unverifiable assumptions about the choice set formation process in order to achieve point identification. Our primary contribution, therefore, is that we offer a new, robust, informative, and implementable method of discrete choice analysis when choice sets are unobserved. We show how one can use this method to partially identify and conduct inference on the distribution of preferences as well as the distribution of choice set size (with an additional independence assumption). We also contribute a new, robust approach to testing assumptions on choice set formation when the setting features suboptimal choices.

In addition to our contributions to the discrete choice literature, our empirical application contributes new insights to the literature on risky choice. In particular, one of our key empirical findings is that our data can be explained by expected utility theory with lower levels of risk aversion than would be implied by many familiar models in the literature. As noted above, the risky choice literature, motivated in part by advances in behavioral economics including the Rabin (2000) critique, has increasingly focused on models that depart from expected utility theory in their specification of *how* agents evaluate risky alternatives. While these models are important and yield many valuable insights, our findings highlight the importance and promise of models that differ in their specification of *which* alternatives agents evaluate. They also highlight the need for and value of data collection efforts that seek to directly measure agents' heterogeneous choice sets.

References

Abaluck, J. and A. Adams (2020): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," *Quarterly Journal of Economics*, forthcoming.

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

Apesteguia, J. and M. A. Ballester (2018): "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences," *Journal of Political Economy*, 126, 74–106.

Artstein, Z. (1983): "Distributions of Random Sets and Random Selections," *Israel Journal of Mathematics*, 46, 313–324.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103, 2499–2529.

——— (2018): "Estimating Risk Preferences in the Field," *Journal of Economic Literature*, 56, 501–564.

Barseghyan, L., F. Molinari, and J. C. Teitelbaum (2016): "Inference under Stability of Risk Preferences," *Quantitative Economics*, 7, 367–409.

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. Prince, and J. C. Teitelbaum (2011): "Are Risk Preferences Stable Across Contexts? Evidence from Insurance Data," *American Economic Review*, 101, 591–631.

Ben-Akiva, M. and B. Boccara (1995): "Discrete Choice Models with Latent Choice Sets," *International Journal of Marketing Research*, 12, 9–24.

Ben-Akiva, M. E. (1973): "Structure of Passenger Travel Demand Models," Ph.D. Dissertation, Department of Civil Engineering, Massachusetts Institute of Technology.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

Beresteanu, A. and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76, 763–814.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Bhargava, S., G. Loewenstein, and J. Sydnor (2017): "Choose to Lose: Health Plan Choices from a Menu with Dominated Options," *Quarterly Journal of Economics*, 132, 1319–1372.

Bugni, F. A., I. A. Canay, and X. Shi (2015): "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics*, 185, 259–282.

Caplin, A. (2016): "Measuring and Modeling Attention," *Annual Review of Economics*, 8, 379–403.

Caplin, A. and M. Dean (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105, 2183–2203.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): "A Random Attention Model," *Journal of Political Economy*, 128, 2796–2836.

Chamberlin, E. H. (1933): *The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value*, Cambridge, MA: Harvard University Press.

Chesher, A. and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989.

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 4, 157–196.

Chiappori, P.-A., A. Gandhi, B. Salanié, and F. Salanié (2019): "From Aggregate Betting Data to Individual Risk Preferences," *Econometrica*, 87, 1–36.

Cicchetti, C. J. and J. A. Dubin (1994): "A Microeconometric Analysis of Risk Aversion and the Decision to Self-Insure," *Journal of Political Economy*, 102, 169–186.

Cohen, A. and L. Einav (2007): "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, 97, 745–788.

Conlon, C. T. and J. H. Mortimer (2013): "Demand Estimation under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5, 1–30.

Coughlin, M. (2020): "Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D," Working Paper, Department of Economics, Rice University.

Crawford, G. S., R. Griffith, and A. Iaria (2020): "A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity," *Journal of Econometrics*, in press.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): "Inferring Cognitive Heterogeneity from Aggregate Choices," *Econometrica*, 88, 1269–1296.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012): "Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–2980.

Draganska, M. and D. Klapper (2011): "Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data," *Journal of Marketing Research*, 48, 653–669.

Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78, 1264–1298.

Gaynor, M., C. Propper, and S. Seiler (2016): "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service," *American Economic Review*, 106, 3521–57.

Ghosal, S. (2001): "Convergence Rates for Density Estimation with Bernstein Polynomials," *Annals of Statistics*, 29, 1264–1280.

Goeree, M. S. (2008): "Limited Information and Advertising in the U.S. Personal Computer Industry," *Econometrica*, 76, 1017–1074.

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–2682.

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.

Heiss, F., D. McFadden, J. Winter, A. Wuppermann, and B. Zhou (2019): "Inattention and Switching Costs as Sources of Inertia in Medicare Part D," Discussion Paper No. 221, CRC TRR 190.

Ho, K., J. Hogan, and F. Scott Morton (2017): "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program," *RAND Journal of Economics*, 48, 877–905.

Honka, E. and P. Chintagunta (2017): "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36, 21–42.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017): "Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry," *RAND Journal of Economics*, 48, 611–646.

Honka, E., A. Hortaçsu, and M. Wildenbeest (2019): "Empirical Search and Consideration Sets," in *Handbook of the Economics of Marketing,* Volume 1, ed. by J.-P. Dubé and P. Rossi, Amsterdam: Elsevier, 193–257.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017): "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market," *American Economic Journal: Economic Policy*, 9, 192–226.

Johnson, E. J., S. B. Shu, B. G. C. Dellaert, C. Fox, D. G. Goldstein, G. Häubl, R. P. Larrick, J. W. Payne, E. Peters, D. Schkade, B. Wansink, and E. U. Weber (2012): "Beyond Nudges: Tools of a Choice Architecture," *Marketing Letters*, 23, 487–504.

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online Demand under Limited Consumer Search," *Marketing Science*, 29, 1001–1023.

Kimya, M. (2018): "Choice, Consideration Sets, and Atrribute Filters," *American Economic Journal: Microeconomics*, 10, 223–247.

Kohn, M. G., C. F. Manski, and D. S. Mundel (1976): "An Empirical Investigation of Factors Which Influence College-Going Behavior," *Annals of Economic and Social Measurement*, 5, 391–419.

Liang, J. and V. Kanetkar (2006): "Price Endings: Magic and Math," *Journal of Product and Brand Management*, 15, 377–385.

Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When More is Less: Limited Consideration," *Journal of Economic Theory*, 170, 70–85.

Lu, Z. (2019): "Estimating Multinomial Choice Models with Unobserved Choice Sets," Working Paper, Financial Stability Department, Bank of Canada.

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

——— (1977): "The Structure of Random Utility Models," *Theory and Decision*, 8, 229–254.

Manski, C. F. and F. Molinari (2010): "Rounding Probabilistic Expectations in Surveys," *Journal of Business and Economic Statistics*, 28, 219–231.

Manzini, P. and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

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.

Molinari, F. (2020): "Microeconometrics with Partial Identification," in *Handbook of Econometrics*, Volume 7A, Amsterdam: Elsevier, in press.

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292.

Roberts, J. H. and J. M. Lattin (1991): "Development and Testing of a Model of Consideration Set Composition," *Journal of Marketing Research*, 28, 429–440.

Schindler, R. M. and P. N. Kirby (1997): "Patterns of Rightmost Digits Used in Advertising Prices: Implications for Nine-Ending Effects," *Journal of Consumer Research*, 24, 192–201.

Schindler, R. M. and A. R. Wiman (1989): "Effect of Odd Pricing on Price Recall," *Journal of Business Research*, 19, 165–177.

Starc, A. (2014): "Insurer Pricing and Consumer Welfare: Evidence from Medigap," *RAND Journal of Economics*, 45, 198–220.

Stigler, G. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

Swait, J. (2001): "Choice Set Generation Within the Generalized Extreme Value Family of Discrete Choice Models," *Transportation Research Part B*, 35, 643–666.

Sydnor, J. (2010): "(Over)Insuring Modest Risks," *American Economic Journal: Applied Economics*, 2, 177–199.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

Terui, N., M. Ban, and G. M. Allenby (2011): "The Effect of Media Advertising on Brand Consideration and Choice," *Marketing Science*, 30, 74–91.

Thaler, R. H. and C. R. Sunstein (2008): *Nudge: Improving Decisions About Health, Wealth and Happiness*, New Haven: Yale University Press.

Tversky, A. (1972a): "Choice by Elimination," *Journal of Mathematical Psychology*, 9, 341–367.

——— (1972b): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 281–299.

van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010): "Retrieving Unobserved Consideration Sets from Household Panel Data," *Journal of Marketing Research*, 47, 63–74.

Vanhuele, M. and X. Drèze (2000): "Do Consumers Really Know if the Price Is Right? Direct Measures of Reference Price and Their Implications For Retailing," Research Paper No. 711, HEC Paris.

Weitzman, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641–654.

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

September 21, 2020

A Theory

A.1 Unobserved Heterogeneity in Choice Sets as Additively Separable Errors

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

A.2 Random Closed Sets

The theory of random closed sets generally applies to the space of closed subsets of a locally compact Hausdorff second countable topological space $\mathbb{F}$. For simplicity we consider here the case $\mathbb{F} = \mathbb{R}^k$ and refer to Molchanov (2017) for the general case. Denote by $\mathcal{F}$ (respectively, $\mathcal{K}$) the collection of closed (compact) subsets of $\mathbb{R}^k$. Denote by $(\Omega, \mathfrak{F}, P)$ the nonatomic probability space on which all random variables and random sets are defined.

Definition A.1 (Random Closed Set): *A map $Y : \Omega \to \mathcal{F}$ is a* random closed set *if for every compact set K in $\mathbb{R}^k$, $Y^{-1}(K) = \{\omega \in \Omega : Y(\omega) \cap K \neq \varnothing\} \in \mathfrak{F}$.*

Definition A.2 (Selection): *For any random set Y, a (measurable)* selection *of Y is a random vector y (taking values in $\mathbb{R}^k$) such that $y(\omega) \in Y(\omega)$, $P-a.s.$*

Theorem A.1 (Artstein's Theorem): *A random vector y and a random set Y can be realized on the same probability space as random elements y' and Y', distributed as y and Y respectively, so that $P(y' \in Y') = 1$, if and only if*

$$P(y \in K) \leqslant P(Y \cap K \neq \varnothing) \ \forall K \in \mathcal{K}. \tag{A.1}$$

Because in this paper the random closed set of interest $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is a subset of $\mathcal{D}$, it suffices to consider $\mathbb{F} = \mathcal{D}$; see Molchanov (2017, Example 1.1.9).

Lemma A.1: *The set $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ in equation (3.1) is a random closed set.*

Proof. Let $D_\kappa^* \equiv D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$. An application of Molchanov (2017, Example 1.1.9) yields that D_κ^* satisfies the measurability requirement in Definition A.1 if the vector $[\mathbf{1}(c \in D_\kappa^*), c \in \mathcal{D}]$ is a random vector with values in $\{0,1\}^{|\mathcal{D}|}$. Next, note that for any $c \in \mathcal{D}$, the event $\{c \in D_\kappa^*\}$ is equivalent to the event $\bigcup_{G \subseteq \mathcal{D}} \{c \in D_\kappa^*, C_i = G\}$. Once the value of C_i is fixed, D_κ^* is a singleton-valued random variable and the result follows. $\square$

A.3 Proof of Theorem 3.1 and Related Results

A.3.1 Proof of Theorem 3.1

Let $d^*(G; \mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ denote the model implied optimal choice for an agent with attributes $(\mathbf{x}, \boldsymbol{\nu})$ and choice set G. Recall that by Assumption 2.2(II), $\Pr(C = G|\mathbf{x}, \boldsymbol{\nu}) = 0$ for all $G \subseteq \mathcal{D}$ such that $|G| < \kappa$. Then by definition the sharp identification region Θ_I is given by the set of values of $\boldsymbol{\theta}$ for which there exists a distribution $\mathsf{F}(\cdot; \mathbf{x}, \boldsymbol{\nu})$ such that $\mathsf{F}(G; \mathbf{x}, \boldsymbol{\nu}) \geqslant 0$ for all $G \subseteq \mathcal{D}$, $\mathsf{F}(G; \mathbf{x}, \boldsymbol{\nu}) = 0$ if $|G| < \kappa$, $\sum_{G \subseteq \mathcal{D}} \mathsf{F}(G; \mathbf{x}, \boldsymbol{\nu}) = 1$, and for all $c \in \mathcal{D}$

$$\Pr(d = c|\mathbf{x}_i) = \int_{\boldsymbol{\tau} \in \mathcal{V}} \sum_{G \subseteq \mathcal{D}} \mathbf{1}(d^*(G; \mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta}) = c) \mathsf{F}(G; \mathbf{x}, \boldsymbol{\tau}) dP(\boldsymbol{\tau}; \gamma), \ \mathbf{x} - a.s. \tag{A.2}$$

This is because for such values of $\boldsymbol{\theta}$ one can complete the model with a distribution $\mathsf{F}(\cdot; \mathbf{x}, \boldsymbol{\nu})$ so that the model implied conditional distribution of optimal choices matches the distribution of choices observed in the data. We are then left to show that this set is equal to the one in equation (3.5). Molchanov and Molinari (2018, Theorem 2.33) show that the observed vector $(d, \mathbf{x})$ is a selection of the random closed set $(D_\kappa^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}), \mathbf{x})$ if and only if the condition in equation (3.5) holds $\mathbf{x}-$a.s. for all $K \subseteq \mathcal{D}$. Take $\boldsymbol{\theta}$ such that there exists a distribution $\mathsf{F}(G; \mathbf{x}, \boldsymbol{\nu})$ under which equation (A.2) holds. By definition $(d^*(G; \mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}), \mathbf{x})$ is a selection of $(D_\kappa^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}), \mathbf{x})$, and by Molchanov and Molinari (2018, Theorem 2.33) the inequality in equation (3.5) holds $\mathbf{x}-$a.s. for all $K \subseteq \mathcal{D}$. Conversely, take a value of $\boldsymbol{\theta}$ for which the inequalities in equation (3.5) are satisfied $\mathbf{x}-$a.s. for all $K \subseteq \mathcal{D}$. Then, by Theorem A.1, there exists a selection $(\tilde{d}_i(G), \mathbf{x})$ of $(D_\kappa^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}), \mathbf{x})$ such that $\Pr(d = c|\mathbf{x}) = \Pr(\tilde{d}(G) = c|\mathbf{x})$, $\mathbf{x}-$a.s., for all $c \in \mathcal{D}$ for some G such that $|G| \geqslant \kappa$. Let $\mathsf{F}(G; \mathbf{x}, \boldsymbol{\nu})$ equal 1 for one such set G with $\tilde{d}(G) = c$, and equal 0 for all other $G \subseteq \mathcal{D}$. Then equation (A.2) holds $\mathbf{x}-$a.s. for all $c \in \mathcal{D}$. To conclude the proof, we show that if the inequalities in (3.5) hold for all $K \subseteq \mathcal{D} : |K| < \kappa$, then they hold for all $K \subseteq \mathcal{D}$. Recall that the set $D_\kappa^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ comprises the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$. Then any set $K \subseteq \mathcal{D} : |K| \geqslant \kappa$ includes at least the $(|\mathcal{D}| - \kappa + 1)$-th best alternative for all realizations of $\boldsymbol{\nu}$ in $\mathcal{V}$, so that $\Pr(D_\kappa^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \emptyset) = 1$ and the inequality in equation (3.5) holds mechanically. $\square$

A.3.2 An Equivalent Characterization Based on Convex Optimization

We next show that the characterization in Theorem 3.1 can equivalently be written in terms of a convex optimization problem.

Corollary A.1: *Let Assumptions 2.1 and 2.2 hold and let* $\Theta = \Delta \times \Gamma$. *Then*

$$\Theta_I = \left\{ \theta \in \Theta : \max_{\mathbf{u} \in \mathbb{R}^{|\mathcal{D}|} : ||\mathbf{u}|| \leqslant 1} \left[\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \int_{\boldsymbol{\tau} \in \mathcal{V}} \max_{d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta})} \left(\mathbf{u}^\top \mathbf{q}^{d^*} \right) dP(\boldsymbol{\tau}; \boldsymbol{\gamma}) \right] = 0, \mathbf{x} - a.s. \right\}, \tag{A.3}$$

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 (A.3) here.[1] 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

$$\mathbf{u}^\top \mathbf{p}(\mathbf{x}) - \int_{\boldsymbol{\tau} \in \mathcal{V}} \max_{d^* \in D^*_\kappa(\mathbf{x}, \boldsymbol{\tau}; \boldsymbol{\delta})} \mathbf{u}^\top \mathbf{q}^{d^*} dP(\boldsymbol{\tau}; \boldsymbol{\gamma}) \leqslant 0 \tag{A.4}$$

holds for all $\mathbf{u} : ||\mathbf{u}|| \leqslant 1$ if and only if (A.4) 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 $\mathbf{u} \in \mathsf{U}$ 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 (A.4) 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 (A.3) belongs to the set defined in (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 (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 (A.3). □

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, which we henceforth denote $D^1, \ldots, D^h$. Hence, the

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

characterization in (A.3) can be rewritten as

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

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).

A.3.3 Positive Probability of Utility Ties

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

$$D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \bigcup_{G \subseteq \mathcal{D}:|G| \geqslant \kappa} \left\{ \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\} = \bigcup_{G \subseteq \mathcal{D}:|G| = \kappa} \left\{ \arg\max_{c \in G} W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \right\}, \tag{A.5}$$

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 $|\mathcal{D}| - \kappa + 1$.[2]

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 $d_i \in D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, $\mathbf{x}_i$–a.s. 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).

[2] 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 the cardinality of $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ equals 3.

A.4 Computational Simplifications

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

A.4.1 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 A.2: *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 $\forall \nu \in \mathcal{V}$ an element of K (possibly different across values of ν) is among the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, then set $\mathbb{K} = \mathbb{K} \backslash K$;[3]

(q) Repeat the following step for $q = 2, \ldots, \kappa - 1$. 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 $\nexists \nu \in \mathcal{V}$ 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, $\Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing) = 1$. Step (q) follows because under the stated condition, the events $\{D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap \{c_j\} \neq \varnothing\}$ and $\{D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K_{q-1} \neq \varnothing\}$ 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 A.2 can be further simplified. The following corollary provides an example.

Corollary A.2: *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 $\{c_j, c_{j+1}, \ldots, c_{j+|\mathcal{D}|-\kappa}\}$, for $j = 1, \ldots, \kappa$.*

[3] 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 A.2 can be initialized to*

$$\mathbb{K} = \Big\{\{c_1\}, \{c_1, c_2\}, \{c_1, c_2, c_3\}, \cdots, \{c_1, c_2, \ldots, c_{\kappa-1}\},$$
$$\{c_{|\mathcal{D}|}\}, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}\}, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}, c_{|\mathcal{D}|-2}\}, \cdots, \{c_{|\mathcal{D}|}, c_{|\mathcal{D}|-1}, \ldots, c_{|\mathcal{D}|-\kappa+2}\}\Big\} \quad \text{(A.6)}$$

and it includes $2(\kappa - 1)$ *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

$$D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \backslash (K_1 \cup K_2) \neq \varnothing \Rightarrow D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap (K_1 \cup K_2) \neq \varnothing. \quad \text{(A.7)}$$

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 (A.7).

Finally, we show that it suffices to verify equation (3.5) for the sets $K \in \mathbb{K}$ as specified in equation (A.6). 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 $K^c = \mathcal{D} \backslash K$,

$$\begin{aligned} P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}) &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset K^c; \boldsymbol{\gamma}) \\ &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_1, \ldots, c_{p-1}\}; \boldsymbol{\gamma}) - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_{q+1}, \ldots, c_{\mathcal{D}}\}; \boldsymbol{\gamma}) \\ &= 1 - P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset \{c_{q+1}, \ldots, c_{\mathcal{D}}\}; \boldsymbol{\gamma}), \end{aligned} \quad \text{(A.8)}$$

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 (A.8) 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 (A.8) 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 $K = \{c_q, \ldots, c_{|\mathcal{D}|}\}$. □

When Assumption 3.1 is maintained, the logic of Theorem A.2 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 A.2 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 A.2 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 A.2 that we use in our application in Sections 4–5. Proofs are omitted because the corollaries follow immediately from Theorem A.2.

Corollary A.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 ν is a scalar 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 $\mathbb{K}=\{K : K \subsetneq \mathcal{D}\}$. 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 A.4: *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 ν is a scalar 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 $\mathbb{K}=\{K : K \subsetneq \mathcal{D}\}$. 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 Theorem A.2 and Corollaries A.3–A.4 is 390 for the sharp identification region of $\boldsymbol{\gamma}$; 1,105 for the sharp identification region of π_5; and 975 for the sharp identification region of π_4.

A.4.2 Additively Separable Extreme Value Type 1 Unobserved Heterogeneity

We now explain how to compute $P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})\cap K\neq\emptyset;\boldsymbol{\gamma})$ 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 $c'\in G$ (and no utility ties), we have

$$\begin{aligned}\Pr(d^*(G;\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})=c'|\mathbf{x},\boldsymbol{\upsilon}) &= \Pr(W(\mathbf{x}_{c'},\boldsymbol{\nu};\boldsymbol{\delta})\geqslant W(\mathbf{x}_c,\boldsymbol{\nu};\boldsymbol{\delta})\ \forall c\in G|\boldsymbol{\upsilon})\\ &= \frac{\exp(\omega(\mathbf{x}_{c'},\boldsymbol{\upsilon};\boldsymbol{\delta}))}{\sum_{c\in G}\exp(\omega(\mathbf{x}_c,\boldsymbol{\upsilon};\boldsymbol{\delta}))}. \end{aligned} \tag{A.9}$$

We show that conditional on $\boldsymbol{\upsilon}$, one can continue to leverage the closed form expressions in equation (A.9) to compute $P(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})\cap K\neq\emptyset;\boldsymbol{\gamma})$ so that numerical integration is needed *only* for $\boldsymbol{\upsilon}$.

Theorem A.3: *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* $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \emptyset | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ *can be computed as a linear combination over different sets* G *of expression* (A.9). *Hence, any* $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \emptyset; \boldsymbol{\gamma})$ *can be computed as an integral with respect to the distribution of* $\boldsymbol{\upsilon}$ *of linear combinations over different sets* G *of expression* (A.9).

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 A.1: *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*

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

Proof of Claim A.1. We first establish equivalence of the following events:

$$\begin{aligned} &\{\exists c' \in K\ s.t.\ W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta});\ \forall c \in \mathcal{D}\backslash K\} \\ &\qquad \iff \cup_{c' \in K}\{W(\mathbf{x}_{c'}, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta}), \forall c \in \mathcal{D}\backslash c'\}. \end{aligned} \tag{A.10}$$

The right-to-left implication in (A.10) 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 (A.10) 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 (A.10) are disjoint. □

Next, recall that, as discussed in Section A.3.2, the set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ can only take on a finite number of realizations, denoted $D^1, \ldots, D^h$, with $|D^j| = |\mathcal{D}| - \kappa + 1$ for all $j = 1, \ldots, h$. We show how to compute the probability of any of these realizations.

Claim A.2: *For each* $j = 1, \ldots, h$, $P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) = D^j | \boldsymbol{\upsilon}; \boldsymbol{\gamma})$ *can be computed as a linear combination of expression* (A.9) *for different sets* G.

Proof of Claim A.2. 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

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

Each of the first two terms in this expression is obtained applying expression (A.9) with $G = \mathcal{D}\backslash D^j$; the last term, by Claim A.1, is obtained as the sum over $c' \in D^j$ of expression (A.9) 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 A.1. □

With these results in hand, we prove Theorem A.3.

Proof of Theorem A.3. By Claim A.2 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 (A.9) with different sets G. To obtain the result in Theorem A.3, 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 $D^j \cap K \neq \varnothing$. □

B Statistical Inference

Under Theorem 3.1, $\boldsymbol{\theta} = (\gamma_1, \gamma_2)$. Under Corollary 3.1, $\boldsymbol{\theta} = (\gamma_1, \gamma_2, \pi_3, \pi_4, \pi_5)$. The sample moments that we use to make confidence statements on (projections of) $\boldsymbol{\theta}$ in Section 5 are:

$$\begin{aligned}
\bar{m}_{n,K,j}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} m_{K,j}(d_i,\mu_i,\mathbf{p}_i;\boldsymbol{\theta})\\
&\equiv \frac{1}{n}\sum_{i=1}^{n}\left[(\mathbf{1}(d_i \in K, (\mu_i,\mathbf{p}_i) \in B_j) - P(D^*_\kappa(\mu_i,\mathbf{p}_i) \cap K \neq \varnothing;\boldsymbol{\gamma})\mathbf{1}((\mu_i,\mathbf{p}_i) \in B_j)\right], \quad \text{(B.1)}
\end{aligned}$$

where B_j, $j = 1,\ldots,J$, are "hypercubes" as defined in Andrews and Shi (2013, Example 1) [AS henceforth], to which we return below, and $P(D^*_\kappa(\mu,\mathbf{p}) \cap K \neq \varnothing;\boldsymbol{\gamma})$ is a function known up to $\boldsymbol{\theta}$ that can be evaluated using the Beta cumulative distribution function.

We obtain confidence regions for the vector $\boldsymbol{\theta}$ using the procedure proposed by AS, as for example in Figure 5.1, and confidence intervals for single components and scalar smooth functions of $\boldsymbol{\theta}$ using the procedure proposed by Kaido et al. (2019), as for example in Table 5.1. Here we briefly recap these procedures. We refer to the original papers for a thorough discussion of the methods, and to Canay and Shaikh (2017) for a comprehensive presentation of the literature on inference in moment inequality models.

Given how the company generates $\mathbf{p}_i$, a household's base price $\bar{p}_i$ is a sufficient statistic for $\mathbf{p}_i$. We follow AS and transform $(\mu_i, \bar{p}_i)$ using the upper-triangular Cholesky decomposition of their sample covariance matrix, so that the transformed variables $(\tilde{\mu}_i, \tilde{p}_i)$ have a sample covariance matrix equal to the identity matrix. We then let the side lengths of the hypercubes B_j be determined by the octiles of the distribution of $\tilde{\mu}_i$ and of the distribution of $\tilde{p}_i$, and include also a hypercube to which all values of $(\mu_i, \mathbf{p}_i)$ belong, so that $J = 65$. As the sample size goes to infinity, the collection of hyper-cubes is required to expand as discussed in AS (p. 624 and Appendix B of their Supplemental Material).

We base our confidence sets on the Kolmogorov-Smirnov test statistic suggested by AS (equation (3.7) on p. 618), which in our framework simplifies to

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

with $\hat{\sigma}_{n,K,j}(\boldsymbol{\theta})$ the sample analog estimator of the population standard deviation of $m_{K,j}(d_i, \mu_i, \mathbf{p}_i; \boldsymbol{\theta})$. Our application of the method proposed by AS computes bootstrap-based critical values to obtain a confidence set

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

where $\eta > 0$ is an arbitrarily small positive constant which we set equal to 10^{-6} as suggested by AS (p. 625). This confidence set covers each $\boldsymbol{\theta} \in \Theta_I$ with asymptotic probability $1-\alpha$ uniformly over a large class of probability distributions $\mathcal{P}$; for a formal statement see AS (Theorem 2 on p. 632). We use this method to compute a confidence set on $\boldsymbol{\gamma} = [\gamma_1, \gamma_2] \in \Gamma \subset \mathbb{R}^2$, and from that to obtain a confidence set on $(\mathrm{E}(\boldsymbol{\nu}), \mathrm{Var}(\boldsymbol{\nu}))$, leveraging the fact that for $\boldsymbol{\nu} \sim Beta(\gamma_1, \gamma_2)$, to each value of (γ_1, γ_2) corresponds a unique pair $(\mathrm{E}(\boldsymbol{\nu}), \mathrm{Var}(\boldsymbol{\nu}))$.

In practice, we evaluate $T_n(\boldsymbol{\theta})$ and the bootstrap-based critical value $\hat{c}_{n,1-\alpha+\eta}(\boldsymbol{\theta})$ on a grid of values $\boldsymbol{\theta}$ designed to give good coverage of the $(\mathrm{E}(\boldsymbol{\nu}), \mathrm{Var}(\boldsymbol{\nu}))$ 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 $\boldsymbol{\nu}$ has a Beta distribution with support $[0, 0.03]$, $\mathrm{E}(\boldsymbol{\nu}) \in 0.03 \times (0,1]$ and $\mathrm{Var}(\boldsymbol{\nu}) \in 0.0009 \times (0, 0.25]$. We therefore obtain a grid of values over (γ_1, γ_2) comprised of 665,603 points, such that the associated grid on $(\mathrm{E}(\boldsymbol{\nu}), \mathrm{Var}(\boldsymbol{\nu}))$ 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. The approximation of $\hat{c}_{n,1-\alpha+\eta}(\boldsymbol{\theta})$ is based on the bootstrap procedure detailed in AS (Section 9) and uses 1,000 bootstrap replications.[4] The procedure takes as inputs a *GMS function* φ, a *GMS sequence*

[4] Compared to the description in AS (Section 9), note that our moment inequalities are of the $\leqslant$ form,

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):[5]

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

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).

We obtain confidence intervals on π_3, π_4, π_5, $\mathrm{E}(\nu)$, and $\mathrm{Var}(\nu)$ using the method proposed by Kaido et al. (2019) [KMS henceforth]. The first three parameters are linear projections of $\boldsymbol{\theta} = [\boldsymbol{\pi}, \boldsymbol{\gamma}]$. The other two are smooth functions of $\boldsymbol{\gamma}$ with gradients that satisfy the assumptions in KMS (Theorem 3.1 on p. 1407). To keep a compact notation, in what follows we denote any function of $\boldsymbol{\theta}$ for which we compute a confidence interval as $f(\boldsymbol{\theta})$. The lower and upper points of the confidence interval (henceforth, CI_n^f) are obtained solving, respectively,

$$\min_{\boldsymbol{\theta}\in\Theta} / \max_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{\theta}) \text{ s.t. } \sqrt{n}\bar{m}_{n,K,j}(\boldsymbol{\theta})/\hat{\sigma}_{n,K,j}(\boldsymbol{\theta}) \leqslant \hat{c}_n^f(\boldsymbol{\theta}),\ j = 1, ..., J,\ K \in \mathbb{K},$$

where $\hat{c}_n^f(\boldsymbol{\theta})$ is computed using the bootstrap-based calibrated projection procedure detailed in KMS (Section 2.2). The critical level $\hat{c}_n^f(\boldsymbol{\theta})$ is calibrated so that the (scalar-valued) function of $\boldsymbol{\theta}$, rather than the vector $\boldsymbol{\theta}$ itself as in AS, is uniformly asymptotically covered with probability $1 - \alpha$ over a large class of probability distributions $\mathcal{P}$, see KMS (Theorem 3.1 on p. 1407) for a formal statement. Similarly to AS, the KMS procedure takes as inputs a GMS function φ and a GMS sequence τ_n.[6] To simplify computations, we use the *hard threshold* GMS function[7]

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

The procedure also requires a regularization parameter $\rho \geqslant 0$, which (similarly to φ and τ_n) enters the calibration of $\hat{c}_{n,1-\alpha}^f$ and introduces a conservative distortion that is required

whereas AS's are of the $\geqslant$ form.

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

[6]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.

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

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.

C Additional Results

C.1 Claim Probabilities

As we explain in Section 4.2, 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.

C.2 Deductible Choices

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

Table C.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).

Table C.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	0.00043	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

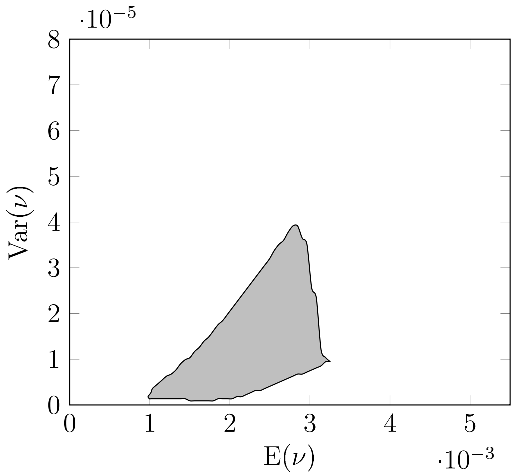
Notes: 95 percent confidence intervals. LB = lower bound. UB = upper bound. Implied risk premia for a lottery that yields a loss of \$1000 with probability 10 percent.

C.3 Subgroup Results

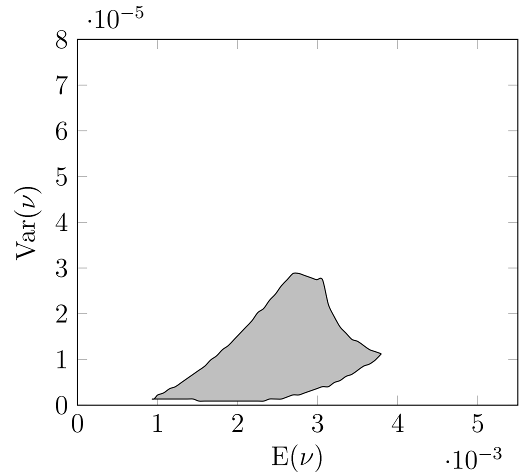
Figure C.1 depicts the AS 95 percent confidence set for $(\mathrm{E}(\nu_i), \mathrm{Var}(\nu_i))$ for population subgroups based on gender, age, and insurance score of the principal driver. In addition, Table C.2 reports (i) the KMS 95 percent confidence interval for the mean of ν_i and (ii) 95 percent 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 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).[8] Strikingly, the lower bounds on the 75th percentile for these two subgroups correspond to risk premia of 17 cents and \$7, respectively.

Table C.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. By implication, one minus the former is the minimum fraction of households who require full-1 or full-2 choice sets to rationalize their deductible choices, while one minus the latter (which equals the lower bound

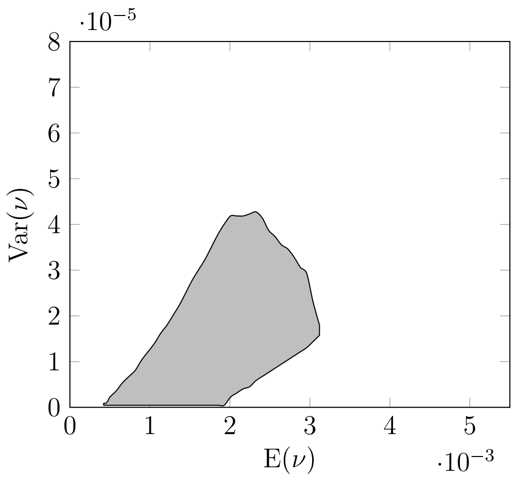
[8]Because the subgroups all have different confidence sets (as well as different sample sizes), it is possible that a result for all households is not a weighted average of the corresponding results within a subgroup.



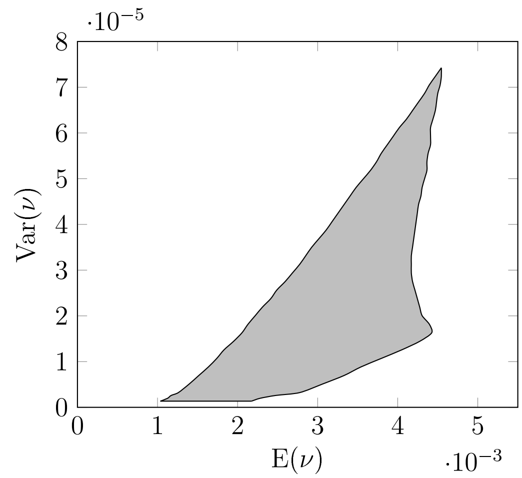
(a) Male



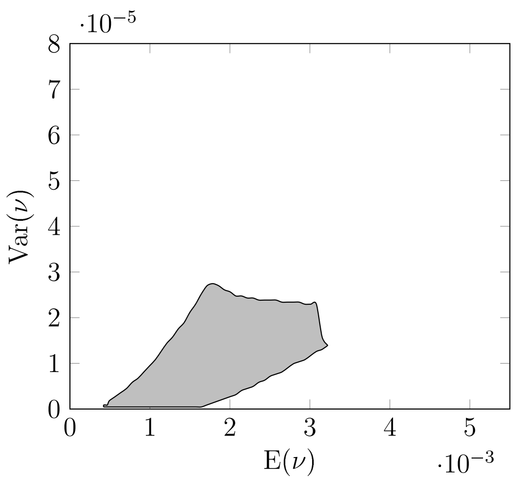
(b) Female



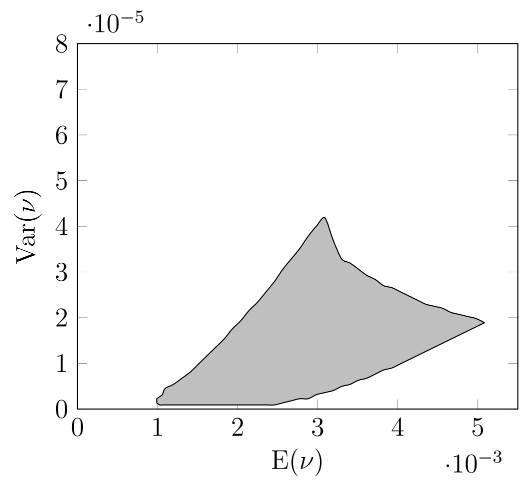
(c) Young



(d) Old



(e) Low insurance score



(f) High insurance score

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

Table C.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.

on π_3) is the minimum fraction of households who require full-2 choice sets.[9] 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 male 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.[10]

C.4 Admissible Probability Density Functions

Figure C.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

[9] By construction, because $\kappa = 3$, the lower bounds on π_5 and π_4 are zero, the lower bound on π_3 is one minus the upper bound on π_4, and the upper bound on π_3 is one.

[10] 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).

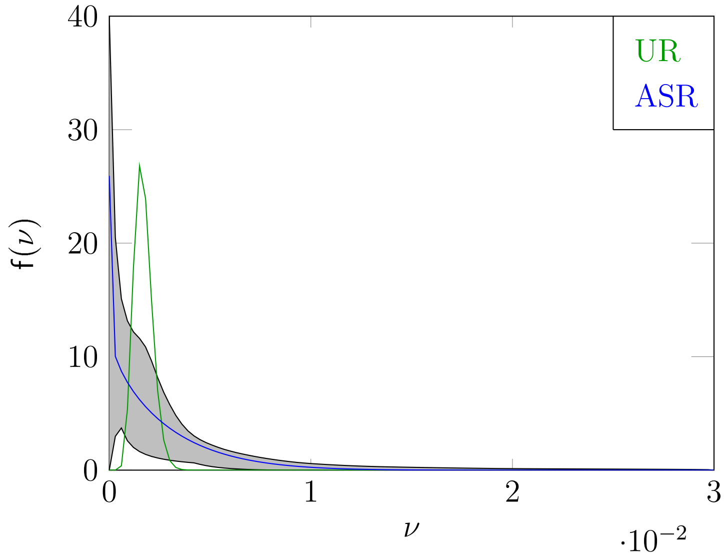


Figure C.2: Confidence set for outer region of admissible probability density functions of ν.

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.

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).

C.5 Proof of Claim 5.1

Claim 5.1(I) follows from Property 5.1 by integrating with respect to the distribution of $\boldsymbol{\nu}$.

Claim 5.1(II) follows from the fact that the UR model satisfies Property 5.1. 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 5.1(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 best alternative in $\mathcal{D}\backslash\{a, b\}$ for some positive measure of values $\boldsymbol{\nu} \in \mathcal{V}$.

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.

Canay, I. A. and A. M. Shaikh (2017): "Practical and Theoretical Advances in Inference for Partially Identified Models," in *Advances in Economics and Econometrics: Eleventh World Congress,* Vol. 2, ed. by B. Honóre, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge: Cambridge University Press, 271–306.

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

Kaido, H., F. Molinari, and J. Stoye (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432.

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

Molchanov, I. (2017): *Theory of Random Sets* (Second Ed.), London: Springer.

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