Heterogeneous Choice Sets and Preferences*

Levon Barseghyan
Cornell University
lb247@cornell.edu

Maura Coughlin
Cornell University
mc2463@cornell.edu

Francesca Molinari
Cornell University
fm72@cornell.edu

Joshua C. Teitelbaum
Georgetown University
jct48@georgetown.edu

June 30, 2019

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 establish that the model is partially identified and characterize its sharp identification region. We also show how the model can be used to assess the welfare cost of limited choice sets. 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 relatively low levels of risk aversion and heterogeneous choice sets. We also find that a mixed logit model, as well as some familiar models of choice set formation, are rejected in our data.

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

*We are grateful to Panle Jia Barwick, Aureo de Paula, Jean-Francois Houde, Chuck Manski, Ulrich Müller, Matthew Thirkettle, Lin Xu, and seminar and conference participants at the Third Annual 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, Caltech, Cornell (Behavioral and Econometrics), Duke, Leuven, Microsoft, Penn, Princeton (Behavioral and Econometrics), Rice, Tilbergen, UCL, UNC, USC, UVA, 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 showed 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 showed 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 simply 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 is cast into doubt 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 there is unobserved heterogeneity in choice sets. Our core model, which imposes mild restrictions on agents' preferences, assumes nothing about agents' choice sets or how they are formed, apart from assuming that they have a known minimum size greater than one. In our main theoretical result, we establish that the distribution of preferences is partially identified and characterize its sharp identification region. The fact that the identification region is sharp implies that it describes all and only those preference distributions for which there exists a choice set distribution such that the model implied distribution of choices matches the distribution of observed choices. It therefore can be used to construct a critical region for rejecting any hypothesized choice set formation process (in conjunction with the model of preferences). 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 the latter is also partially identified. In addition, we show how one can use our approach to assess the welfare cost of limited choice sets (i.e., choice sets that do not contain all feasible alternatives).

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 constant greater than one. Consequently, our model admits a wide range of possible choice set formation processes 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 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.

In the remainder of the section we prove three identification results. First, we show that under the minimal assumptions of our core model, the distribution of preferences is partially identified, without the need for additional assumptions about choice sets or how they are formed. Second, we show that with one additional restriction on the choice set formation process—namely, that choice set size is independent of preferences—the distribution of choice set size is also partially identified. In both cases, we 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 of the model's parameters. Lastly, we characterize the sharp upper bound on the welfare cost of limited choice sets as the solution to a maximization problem whose objective is a smooth function of the core model's parameters.

In the two ensuing sections 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. We also apply our findings to assess the welfare cost of limited choice sets in this context. The data hail from a large U.S. insurance company and contain information on more than 100,000 households who first purchased auto policies 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 aversion and choice sets and that fits the random utility model framework that we develop in the two preceding sections. Our empirical model assumes, inter alia, that households have expected utility preferences and exhibit constant absolute risk aversion, that their risk aversion conditional on observables follows a Beta distribution, and that their choice sets contain at least three alternatives. After specifying the model, we describe our data and present sample statistics.

We then discuss how certain patterns in the data—which relate to the fact that a sizable fraction of households choose a suboptimal alternative—are suggestive of unobserved heterogeneity in choice sets and cannot be explained by standard discrete choice models (e.g., mixed logit). We also discuss how these patterns are consistent with some models of heterogeneous choice set formation, but not others. The import of this discussion goes beyond our specific application and contributes new testable implications for any random utility model that fits our framework and is applied to a context in which the feasible set contains suboptimal alternatives under the model. As we demonstrate within the context of our application, one can leverage these implications to test the model's assumptions on the choice set formation process under weak restrictions on the utility function and without functional form restrictions on the distribution of preferences or unobservables.

We present our empirical findings in Section 5. To start, we employ the generalized moment selection procedure of Andrews and Soares (2010) to obtain a 95 percent confidence set for the parameters of the risk preference model. Parameter values inside the confidence set describe the distributions of risk preferences for which there exists a distribution of choice sets (of minimum size three) such that the distribution of choices implied by the model matches the distribution of choices observed in the data. Accordingly, parameter values outside the confidence set are rejected, and so are the models that—when estimated—yield such values. In our application the confidence set proves to be highly informative. For instance, we find that the distribution of risk preferences estimated by a mixed logit model with full sized choice sets (i.e., choice sets that contain all feasible alternatives) is rejected, as is the distribution estimated by our empirical model when coupled with the assumption that choice sets are drawn uniformly at random from the feasible set conditional on their size and independently of preferences (cf. Dardanoni et al. 2018). By contrast, we find that the distribution of risk preferences estimated by our empirical model when coupled with one variant of the assumption that feasible alternatives independently enter the choice set with alternative specific probabilities and independently of preferences (cf. Manski 1977; Manzini and Mariotti 2014) is not rejected. It is important to note that the rejections described in this paragraph are different from and not necessarily implied by the rejections described in the previous paragraph. A rejection there is a rejection of a specific choice set formation process combined with any distribution of preferences in a given class of utility models. A rejection here is a rejection of a specific distribution of risk preferences combined with any choice set formation process (subject to the minimum size restriction).

Next, we apply the calibrated projection method of Kaido et al. (2019) to obtain 95 percent confidence intervals for selected smooth functions and projections of the model's parameters, including moments of the distribution of risk aversion, the maximum welfare cost of limited choice sets, and the distribution of choice set size. Our key finding with respect to risk aversion is that our estimated lower bounds are substantially smaller than the point estimates obtained under several comparator models (including those we obtain using a mixed logit model with full sized choice sets and those obtained by Cohen and Einav (2007) using a Poisson-Gaussian mixture model with full sized choice sets). This suggests that the data can be explained by expected utility theory with lower and more homogeneous levels of risk aversion than would be implied by many familiar models in the literature. We also find that the welfare cost of limited choice sets may be as high as 25 percent of what the average household spends on auto collision coverage, and that at least 80 percent of households require limited choice sets to explain their deductible choices.

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 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 is an econometric representation of utility theory in which the utility function is a random variable. The model posits that for each agent $i \in \mathcal{I}$ with *choice set* $C_i \subseteq \mathcal{D}$ 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 we assume the probability of ties is zero.

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 a convex compact parameter space, 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 probability 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 a convex compact parameter space, and independent of* $\mathbf{x}_i$*.*

Assumption 2.1(I) restricts the family $\mathcal{U}$ from which the utility function U_i is drawn to be a known parametric class. It is weaker than the assumption, typically imposed in discrete choice models, that U_i is additively separable in unobservables. Assumption 2.1(II) allows for agent specific unobserved heterogeneity in U_i, indexed by the vector $\boldsymbol{\nu}_i$. It restricts the distributional family of $\boldsymbol{\nu}_i$ to be a known parametric class. It also requires that $\boldsymbol{\nu}_i$ is independent of $\mathbf{x}_i$, though one can relax this restriction based on the specific structure of the empirical model (as we illustrate in our application). These restrictions are in line with the distributional assumptions in standard discrete choice models, such as the conditional logit model of McFadden (1974) and the mixed logit model of McFadden and Train (2000).[1] However, we emphasize that the parametric restrictions on W and P are not essential for our partial identification results in Section 3; see Remark 3.1.

Remark 2.1: Due to the ordinal nature of the model, the family $\{W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) : \boldsymbol{\delta} \in \Delta\}$ cannot include two functions that are monotone transformations of each another (Matzkin 2007). Also, to ensure the probability of ties is zero, the functions W and P must satisfy the condition $\Pr(W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = W(\mathbf{x}_{ic'}, \boldsymbol{\nu}_i; \boldsymbol{\delta})) = 0$ for all $c, c' \in \mathcal{D},\ c \neq c'$. We assume that the model satisfies these basic conditions.

Our key point of departure from McFadden (1974) and the bulk of the discrete choice literature lies in 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 (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:

[1] They are stronger, however, than the restrictions in Manski (1975), whose maximum score estimator requires weaker distributional assumptions.

Assumption 2.2 (Random Sample and Empirical Content):

(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 scalar.

Assumption 2.2(I) is weaker than the standard assumption as it omits the requirement that the agents' choice sets, $\{C_i : i \in I \subset \mathcal{I}\}$, are observed. Given this difference, Assumption 2.2(II) is comparable to the standard assumption. Both require that the agents' choice sets contain at least two alternatives, which is necessary for the model to have empirical content. Both also require that the minimum choice set size, which we denote by κ, is known. Under the standard approach this is an implication of the assumption that the agents' choice sets are observed. In Assumption 2.2(II) we assume that κ is known, either from information in the data or by assumption, even though the agents' choice sets are unobserved. In any event, Assumption 2.2(II) is weaker than the assumption, commonly imposed in empirical applications of discrete choice models (though increasingly challenged in the theoretical and empirical literatures), that each agent's choice set coincides with either the feasible set, $C_i = \mathcal{D}$, or a known subset D of the feasible set, $C_i = D \subset \mathcal{D}$.

Remark 2.2: The classic random utility models in the tradition of McFadden (1974), which have the form $U_i(c) = W_i(c) + \epsilon_{ic}$ where ϵ_{ic} is an additive disturbance that is agent and alternative specific, can be subsumed within our framework; see Appendix A.

The random utility model presented in this section admits a wide range of choice set formation processes that result in unobserved heterogeneity in agents' choice sets. For instance, the model allows for a process in which an agent's choice set C_i is drawn uniformly at random from the feasible set $\mathcal{D}$, conditional on $|C_i| = q$ for $q \geqslant \kappa$ (cf. Dardanoni et al. 2018), or in which each alternative $c \in \mathcal{D}$ enters an agent's choice set C_i with probability $\varphi(c)$ independently of other alternatives, conditional on $|C_i| \geqslant \kappa$ (cf. Manski 1977; Manzini and Mariotti 2014). Importantly, the model allows for any dependence structure, without restriction, (i) between agents' choice sets and their observable attributes and (ii) conditional on observables, between agents' choice sets and their unobservable attributes.

3 Partial Identification of the Model's Parameters

In this section we show that one can partially identify the distribution of preferences without specifying any particular choice set formation process (Section 3.1). We then show that, under an additional restriction on the dependence between choices sets and unobservable attributes, one can also partially identify the distribution of choice set size (Section 3.2). Lastly, we show how one can use our approach to conduct welfare analysis. In particular, we use it to assess the welfare cost of limited choice sets (Section 3.3).

3.1 Preferences

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 model specified in Section 2 implies multiple optimal choices for the agent, resulting from the multiple possible realizations G of her choice set C_i.[2] 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 $\mathcal{D}$.[3] It contains the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$, where "best" is defined with respect to $W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$.

Figure 3.1 is a stylized depiction of the set $D_\kappa^* \equiv D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ for the case where the feasible set is $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$, $\kappa \in \{4, 5\}$, and $\boldsymbol{\nu}_i = \nu_i$ is a scalar. In the figure, $\bar{\nu}_{c_a, c_b}(\mathbf{x}_i)$ is the threshold 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. The construction of D_κ^* is straightforward. Given $(\mathbf{x}_i, \nu_i)$ and $\boldsymbol{\delta}$, rank the alternatives in $\mathcal{D}$ from best to worst according to their utilities. If $\kappa = 5$ the agent draws a choice set of size 5 and hence D_κ^* comprises the first best alternative. If $\kappa = 4$ the agent may draw a choice set of size 4 or 5 and hence D_κ^* 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 of C_i. The agent chooses the first best if it is contained in C_i; otherwise she chooses the second best. For instance, suppose $\nu_i \in (\bar{\nu}_{c_2, c_3}(\mathbf{x}_i), \bar{\nu}_{c_1, c_3}(\mathbf{x}_i)]$. Then $D_\kappa^* = \{c_2, c_3\}$ where c_2 is first best. The agent chooses c_2

[2] This is the case even though there exist different choice sets for which the model implied optimal choice is the same.

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

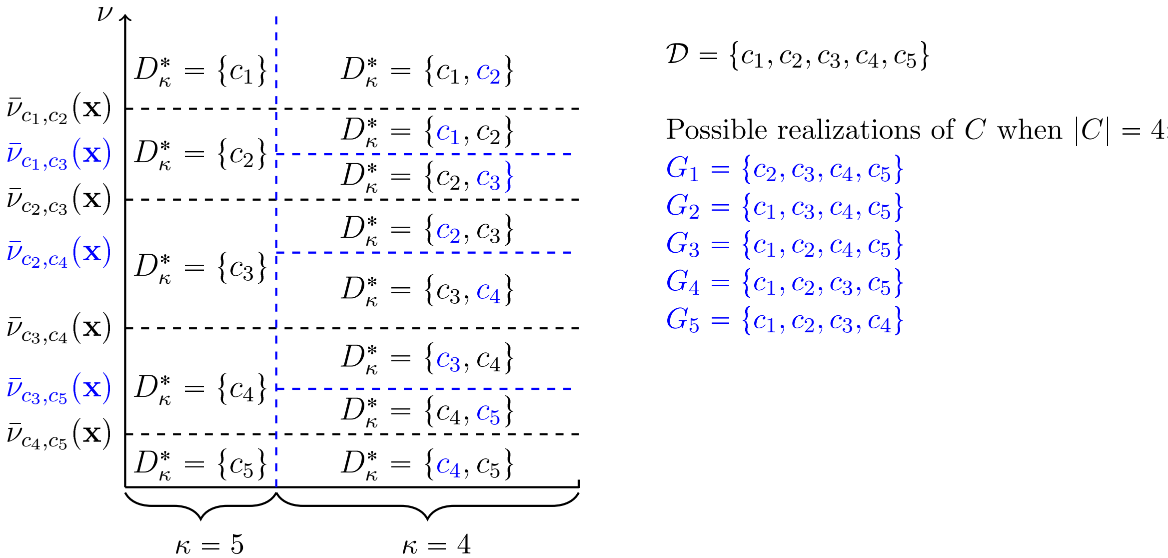


Figure 3.1: Stylized depiction of D^*_κ when $|\mathcal{D}| = 5$ and $\kappa \in \{4, 5\}$.

Note: The figure depicts the set D^*_κ of model implied optimal choices as a function of the agent's unobserved attribute ν and choice set $C = G \subseteq \mathcal{D}$, $|G| \geqslant \kappa$, for $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa \in \{4, 5\}$.

provided that $C_i \neq G_2$; she chooses c_3 only if $C_i = G_2$. More generally, the agent chooses the best alternative in the intersection of D^*_κ and her realization of C_i.

Let $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ denote the conditional probability mass function of C_i given $(\mathbf{x}_i, \boldsymbol{\nu}_i)$. Thus,

$$\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = \Pr(C_i = G | \mathbf{x}_i, \boldsymbol{\nu}_i),$$

where $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) \geqslant 0$ for all $G \subseteq \mathcal{D}$ and $\sum_{G \subseteq \mathcal{D}} \mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = 1$. The model in Section 2 imposes no restrictions on $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$, except to require that $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = 0$ for all $G \subseteq \mathcal{D}$ such that $|G| < \kappa$. In the absence of additional restrictions on $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$, the multiplicity of model implied optimal choices—when it results in overlapping sets $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ of model implied optimal choices for different values of the agent's unobservable attributes $\boldsymbol{\nu}_i$—generally precludes point identification of the model's parameters $\boldsymbol{\theta} = [\boldsymbol{\delta}; \boldsymbol{\gamma}]$.

To see this, let $\Pr(d^*_i = c | \mathbf{x}_i; \boldsymbol{\theta})$ denote the model implied conditional probability that alternative c is chosen given $\mathbf{x}_i$ and $\boldsymbol{\theta}$. Observe that for all $c \in \mathcal{D}$,

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

Imagine that the distribution $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$ was known. Then the parameter vector $\boldsymbol{\theta}$ would be point identified (given sufficient variation in $\mathbf{x}_i$) by the condition that

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

where d_i is the agent's observed choice. When $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$ is unknown, however, there may be multiple combinations of $\boldsymbol{\theta}$ and $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$ that satisfy condition (3.3), due to the multiplicity of model implied optimal choices $d_i^*(G;\mathbf{x}_i,\boldsymbol{\nu}_i;\boldsymbol{\delta})$. The logic is illustrated by Figure 3.1. For an extreme example, suppose we observe that alternative c_3 is always chosen. This is consistent with: (i) $\nu_i \in (\bar{\nu}_{c_2,c_3}(\mathbf{x}_i), \bar{\nu}_{c_1,c_3}(\mathbf{x}_i)]$ and $\mathsf{F}(G_3;\mathbf{x}_i,\boldsymbol{\nu}_i) = 1$; (ii) $\nu_i \in (\bar{\nu}_{c_2,c_4}(\mathbf{x}_i), \bar{\nu}_{c_2,c_3}(\mathbf{x}_i)]$ and $\mathsf{F}(G_3;\mathbf{x}_i,\boldsymbol{\nu}_i) = 0$; (iii) $\nu_i \in (\bar{\nu}_{c_3,c_4}(\mathbf{x}_i), \bar{\nu}_{c_2,c_4}(\mathbf{x}_i)]$ and $\mathsf{F}(G_3;\mathbf{x}_i,\boldsymbol{\nu}_i) = 0$; and (iv) $\nu_i \in (\bar{\nu}_{c_3,c_5}(\mathbf{x}_i), \bar{\nu}_{c_3,c_4}(\mathbf{x}_i)]$ and $\mathsf{F}(G_4;\mathbf{x}_i,\boldsymbol{\nu}_i) = 1$.[4]

The set of values of the parameter vector $\boldsymbol{\theta}$ for which there exists a distribution $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$, satisfying $\mathsf{F}(G;\mathbf{x}_i,\boldsymbol{\nu}_i) = 0$ for all $G \subseteq \mathcal{D}$ such that $|G| < \kappa$, such that condition (3.3) holds forms the *sharp identification region* of $\boldsymbol{\theta}$. We denote this region by Θ_I. The distribution $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$, however, is an infinite dimensional nuisance parameter, which creates difficulties for the computation of Θ_I and for statistical inference.

We circumvent these difficulties by working directly with the set of model implied optimal choices, $D_\kappa^*(\mathbf{x}_i,\boldsymbol{\nu}_i;\boldsymbol{\delta})$. 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) given the distribution of observed choices $(\Pr(d_i = c|\mathbf{x}_i), c \in \mathcal{D})$, $\mathbf{x}_i - a.s.$, we leverage a result in Artstein (1983), reported in Appendix A, Theorem A.1. 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 as the set of values of the parameter vector $\boldsymbol{\theta}$ for which the inequalities hold.[5]

[4] Standard revealed preference arguments presume $\mathsf{F}(\mathcal{D};\mathbf{x}_i,\boldsymbol{\nu}_i) = 1$, i.e., $C_i = \mathcal{D}$. As Figure 3.1 illustrates, however, these arguments break down if C_i is unobserved and $C_i = G \subset \mathcal{D}$ is possible. Indeed, when this is the case, if no restrictions are imposed on $\mathsf{F}(\cdot;\mathbf{x}_i,\boldsymbol{\nu}_i)$, the model is *incomplete* (Tamer 2003) and conventional methods of statistical inference do not apply.

[5] 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 (2019).

THEOREM 3.1: *Let Assumptions 2.1 and 2.2 hold and let* $\Theta = \Delta \times \Gamma$. *Then*

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

It is immediate that if equation (3.4) holds then the inequalities in equation (3.5) are satisfied for each $K \subseteq \mathcal{D}$. We refer to Molchanov (2017, Theorem 1.4.8) for a proof of the fact that if the inequalities in equation (3.5) are satisfied for each $K \subseteq \mathcal{D}$ then equation (3.4) holds. Our proof of Theorem 3.1, provided in Appendix A, establishes that the characterization in equation (3.5) is sharp—all and only those values of $\boldsymbol{\theta}$ for which the inequalities in equation (3.5) hold could have generated the observed data under the maintained assumptions.

For each $K \subset \mathcal{D}$, the left hand side of the inequality in equation (3.5), $\Pr(d_i \in K|\mathbf{x}_i)$, can be estimated from the data $\{(d_i, \mathbf{x}_i) : i \in I \subset \mathcal{I}\}$, and the right hand side is a function of $\mathbf{x}_i$ known up to $\boldsymbol{\theta}$. In Appendix A, Theorem A.2 we provide an algorithm that substantially reduces the number of inequalities that need to be checked to obtain Θ_I.

REMARK 3.1: Theorem 3.1, as well as Corollary 3.1 and Theorem 3.2 below, can be generalized for a *structure* (W, P) (or (W, P, π) in the case of Corollary 3.1) 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.

3.2 Preferences and 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, we now assume that:

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

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

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 a convex compact parameter space. To simply 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 contained in $\{2, \ldots, |\mathcal{D}|\}$, which allows for the possibility that the agent's choice set has full size, $\ell_i = |\mathcal{D}|$, or is limited, $\ell_i < |\mathcal{D}|$. Under this assumption the model continues to admit a wide range of choice set formation processes. 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 and that ℓ_i is independent of $\boldsymbol{\nu}_i$. Conditional on ℓ_i, however, the model continues to allow for any dependence structure, without restriction, (i) between agents' choice sets and their observable attributes and (ii) 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.[6]

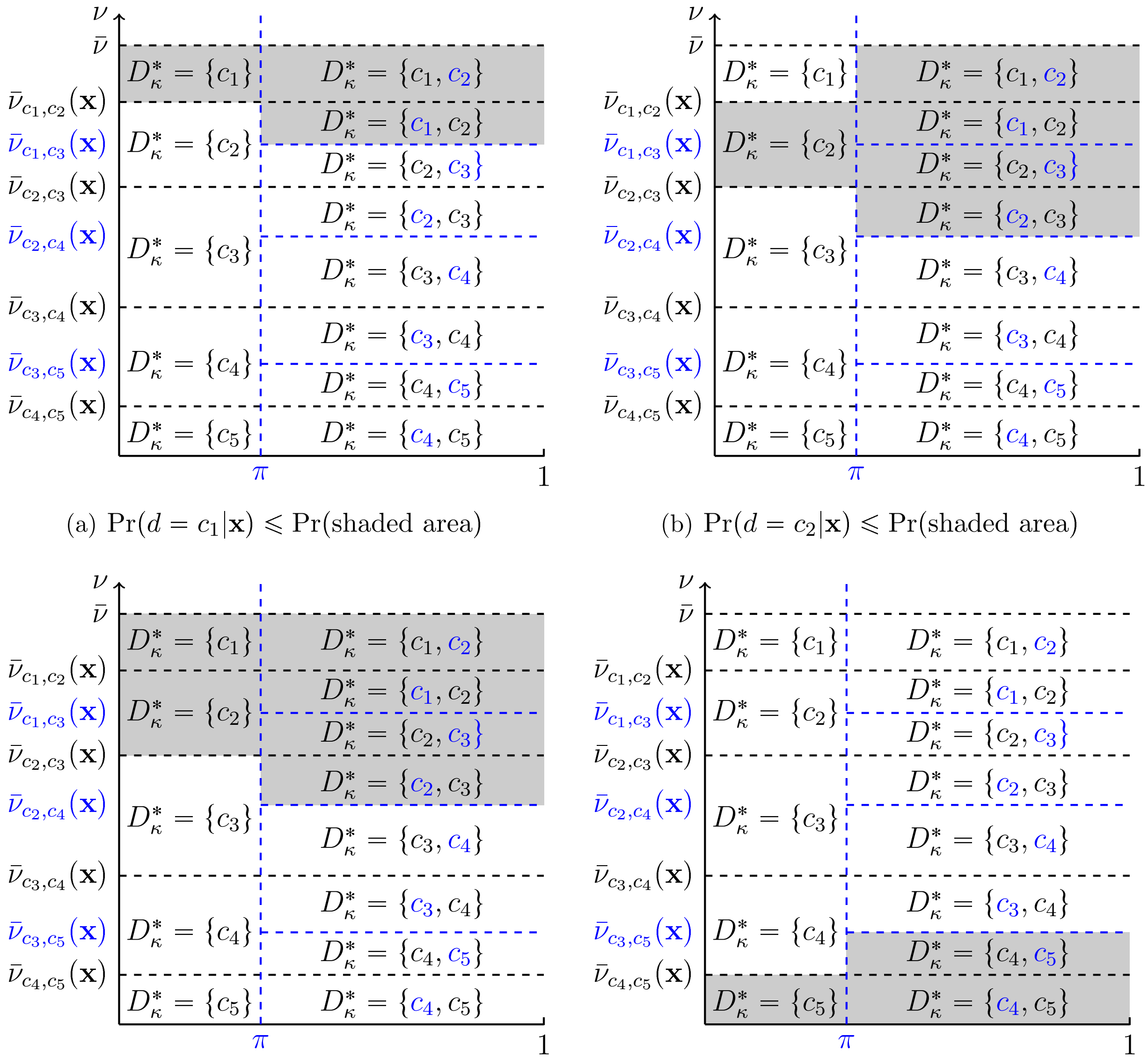
Corollary 3.1: *Let Assumptions 2.1, 2.2, and 3.1 hold and let* $\boldsymbol{\theta} = [\boldsymbol{\eta}; \boldsymbol{\delta}; \boldsymbol{\gamma}]$ *and* $\Theta = H \times \Delta \times \Gamma$. *Then*

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

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 $\kappa < |\mathcal{D}|$), an upper bound on $\pi_{|\mathcal{D}|}(\mathbf{x}_i; \boldsymbol{\eta})$ (the lower bound is zero provided $\kappa < |\mathcal{D}|$ because $D_{q+1}^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \subset D_q^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$), and lower and upper bounds on $\pi_q(\mathbf{x}_i; \boldsymbol{\eta})$ for $q = \kappa + 1, \ldots, |\mathcal{D}| - 1$.

Figure 3.2 contains stylized depictions of selected inequalities in equation (3.7) for the case where $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$, $\kappa = 4$, and $\boldsymbol{\nu}_i = \nu_i$ is a scalar with support on $\mathcal{V} = [0, \bar{\nu}]$. 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 $\ell_i = 5$, in which case D_κ^*

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



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

(d) $\Pr(d = c_5|\mathbf{x}) \leqslant \Pr(\text{shaded area})$

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

Note: The figure depicts the inequalities in equation (3.7) for the following subsets $K \subseteq \mathcal{D}$ when $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 4$: (a) $K = \{c_1\}$; (b) $K = \{c_2\}$; (c) $K = \{c_1, c_2\}$; and (d) $K = \{c_5\}$.

comprises the first best alternative. With probability $1-\pi$ the agent draws a choice set of size $\ell_i = 4$, in which case D_κ^* 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 of C_i. She chooses the first best if it is contained in C_i; otherwise she chooses the second best. As before, $\bar{\nu}_{c_a,c_b}(\mathbf{x}_i)$ is the threshold 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 either $\nu_i > \bar{\nu}_{c_1,c_2}(\mathbf{x}_i)$ and the realization G of C_i includes c_1 or $\nu_i \in (\bar{\nu}_{c_1,c_3}(\mathbf{x}_i), \bar{\nu}_{c_1,c_2}(\mathbf{x}_i)]$ and G excludes c_2. It follows that

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

A similar reasoning applies to the other singleton sets $K = \{c_2\}, \ldots, \{c_5\}$, with $K = \{c_2\}$ depicted in Panel (b).

The inequalities in equation (3.7) also include non-singleton sets $K \subseteq \mathcal{D}$. To see why such inequalities are needed, Panel (c) depicts the case $K = \{c_1, c_2\}$. While

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

by the additivity of probabilities, the right hand side of the inequality is subadditive. As one can see comparing Panels (a) and (b) to Panel (c), the shaded area in Panel (c) is smaller than the sum of the shaded areas in Panels (a) and (b). Hence, values of $\boldsymbol{\theta}$ that satisfy the inequality for $K = \{c_1\}$ and $K = \{c_2\}$ may fail to do so for $K = \{c_1, c_2\}$.

Not all pairs of singleton sets, however, yield non-redundant inequalities. Consider Panel (d), which depicts the inequality for $K = \{c_5\}$. Comparing the shaded area in Panel (d) with that in Panel (a) reveals that c_1 and c_5 cannot occur as multiple optimal choices for the same value of ν_i. In this case, therefore,

$$\Pr(D_\kappa^* \cap \{c_1, c_5\} \neq \varnothing) = \Pr(D_\kappa^* \cap \{c_1\} \neq \varnothing) + \Pr(D_\kappa^* \cap \{c_5\} \neq \varnothing),$$

rendering the inequality for $K = \{c_1, c_5\}$ redundant if the inequalities for $K = \{c_1\}$ and $K = \{c_5\}$ are satisfied. This reasoning can substantially reduce the number of inequalities that are needed to recover Θ_I (as mentioned previously in connection with Theorem 3.1) and is formalized in Appendix A, Theorem A.2.

Though not depicted in Figure 3.2, let us highlight what identifies an upper bound on π. Consider $K = \{c_1, c_2, c_3, c_4\}$. Given this K, we have $\Pr(d_i \in K|\mathbf{x}_i) = 1 - \Pr(d_i = c_5|\mathbf{x}_i)$. At the same time we have $\Pr(D^*_\kappa \cap K \neq \varnothing) = 1 - \Pr(D^*_\kappa = \{c_5\})$. It follows that

$$\Pr(d_i \in K|\mathbf{x}_i) \leqslant \Pr(D^*_\kappa \cap K \neq \varnothing)$$
$$\Leftrightarrow \quad \Pr(d_i = c_5|\mathbf{x}_i) \geqslant \Pr(D^*_\kappa = \{c_5\}) = \pi P(\nu_i \leqslant \bar{\nu}_{c_4,c_5}(\mathbf{x}_i); \boldsymbol{\gamma}).$$

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 $\boldsymbol{\theta}$.

In Sections 4 and 5 we apply Theorem 3.1 and Corollary 3.1 to learn about the distributions of risk preferences and choice set size, respectively, in a model of risky choice with unobserved heterogeneity in preferences and choice sets. We specify our empirical model and discuss our data in Section 4. We present our empirical findings in Section 5. In our application we use the generalized moment selection procedure introduced by Andrews and Soares (2010) to obtain asymptotically uniformly valid confidence sets for $\boldsymbol{\theta}$. We then apply the calibrated projection method proposed by Kaido et al. (2019) to obtain asymptotically uniformly valid confidence intervals for smooth functions and components of $\boldsymbol{\theta}$.[7]

3.3 Welfare Cost of Limited Choice Sets

In the application we also use our approach to assess the welfare cost of limited choice sets. Specifically, we compute the certainty equivalent of the maximum possible gain in model implied expected utility from expanding every household's choice set from minimum size ($\ell_i = \kappa$) to full size ($\ell_i = |\mathcal{D}|$). This provides a measure of the maximum potential welfare cost of limited choice sets. Measuring this cost can be important for market design and public policy. There can be a tradeoff between the cost of ensuring that households draw full sized choice sets and the potential that households make suboptimal choices when their choice sets are limited. The nature of the drivers of limited choice sets—e.g., limited attention or exogenous restrictions—determines whether a market redesign or intervention is worthwhile.

The lower bound on the welfare cost of limited choice sets is zero by definition. It is achieved when, even though $\kappa < |\mathcal{D}|$, every household nevertheless draws a choice set that contains the first best alternative among all feasible alternatives. The upper bound on the welfare cost is the interesting quantity to learn. It is obtained when every household draws a choice set that comprises the κ worst alternatives, in which case the model implies that every household chooses the κth worst alternative. The formal result follows.

[7] For details on our empirical methods, see Section 5.1 and Appendix B.

THEOREM 3.2: *Let Assumptions 2.1 and 2.2 hold and let* $\boldsymbol{\theta} = [\boldsymbol{\delta}; \boldsymbol{\gamma}]$ *and* $\Theta = \Delta \times \Gamma$. *For a given* $\mathbf{x}$ *and realization* $\boldsymbol{\tau}$ *of* $\boldsymbol{\nu}$, *define the ranked (from worst to best) alternatives as*

$$
\begin{aligned}
d^1(\boldsymbol{\tau}) &= \arg\min_{c \in \mathcal{D}} W(\mathbf{x}_c, \boldsymbol{\tau}; \boldsymbol{\delta}) \\
d^2(\boldsymbol{\tau}) &= \arg\min_{c \in \{\mathcal{D} \setminus d^1(\boldsymbol{\tau})\}} W(\mathbf{x}_c, \boldsymbol{\tau}; \boldsymbol{\delta}) \\
d^3(\boldsymbol{\tau}) &= \arg\min_{c \in \{\mathcal{D} \setminus \{d^1(\boldsymbol{\tau}), d^2(\boldsymbol{\tau})\}\}} W(\mathbf{x}_c, \boldsymbol{\tau}; \boldsymbol{\delta}) \\
&\vdots \\
d^{|\mathcal{D}|}(\boldsymbol{\tau}) &= \arg\max_{c \in \mathcal{D}} W(\mathbf{x}_c, \boldsymbol{\tau}; \boldsymbol{\delta}),
\end{aligned}
$$

where the dependence of the ranked alternatives on the choice set, $\mathbf{x}$, *and* $\boldsymbol{\delta}$ *is suppressed to simplify notation. Then the sharp upper bound on the welfare cost of limited choice sets is*

$$
\begin{aligned}
&\max_{\boldsymbol{\theta} \in \Theta} \; \mathrm{E}(W(\mathbf{x}_{d^{|\mathcal{D}|}(\boldsymbol{\nu})}, \boldsymbol{\nu}; \boldsymbol{\delta}); \boldsymbol{\gamma}) - \mathrm{E}(W(\mathbf{x}_{d^{\kappa}(\boldsymbol{\nu})}, \boldsymbol{\nu}; \boldsymbol{\delta}); \boldsymbol{\gamma}) \\
&\text{s.t.} \;\; \Pr(d \in K | \mathbf{x}) \leqslant P(D^*_{\kappa}(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \emptyset; \boldsymbol{\gamma}), \; \forall K \subseteq \mathcal{D}, \; \mathbf{x} - a.s.
\end{aligned}
\tag{3.8}
$$

The proof of Theorem 3.2 follows immediately from Theorem 3.1. Indeed, Theorem 3.1 yields that Θ_I is characterized by the constraints in problem (3.8). For a given $\boldsymbol{\theta} \in \Theta_I$, the gain in model implied expected utility from expanding choice set size from minimum to full is given by the objective function in problem (3.8). Maximizing over Θ_I yields the result.

Under our assumptions, the objective function in problem (3.8) is smooth in the parameter vector $\boldsymbol{\theta}$. Therefore, we can apply the calibrated projection method proposed by Kaido et al. (2019) to obtain an asymptotically uniformly valid confidence interval for the solution to problem (3.8). For details, see Appendix B.

4 Risk Preferences and Choice Sets 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. We also assess the welfare cost of limited choice sets in this context. In Section 4.1 we specify a random expected utility model that allows for unobserved heterogeneity in risk aversion and choice sets and maintains Assumptions 2.1, 2.2, and 3.1. In Section 4.2 we describe our data. In Section 4.3 we discuss patterns in the data that are suggestive of unobserved heterogeneity in choice sets and that cannot be explained by standard discrete choice models. We present our empirical findings in Section 5.

Our application illustrates how one can utilize our approach to learn about agents' heterogeneous preferences from choice data and to conduct welfare analysis when there is or may be unobserved heterogeneity in agents' choice sets. In this spirit, we assume that agents have standard expected utility preferences and make other simplifying assumptions, including constant absolute risk aversion. Our welfare analysis measures the welfare cost of limited choice sets in the worst case scenario where every agent's choice set comprises the κ worst alternatives. We emphasize, however, that one can apply our approach to a wide range of preference models and welfare questions.

4.1 Empirical Model

We model households' deductible choices in auto collision insurance. Each household i faces (i) a menu of prices $\mathbf{p}_i \equiv (p_{ic}, c \in \mathcal{D})$, where p_{ic} is the household specific premium associated with deductible c and $\mathcal{D}$ is the feasible set of deductible options, and (ii) a probability μ_i of experiencing a claim during the policy period. In addition, each household has an array of demographic characteristics $\mathbf{t}_i$.

Following the related literature on property insurance (for a survey, see Barseghyan et al. 2018), we make two simplifying assumptions about claims and their probabilities.

Assumption 4.1 (Claims and Claim Probabilities):

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

(II) Any claim exceeds the highest available deductible; 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.[8] Assumption 4.1(II) abstracts from small claims, transaction costs, and moral hazard. Both assumptions are standard in the literature (e.g., Cohen and Einav 2007; Sydnor 2010; Barseghyan et al. 2011, 2013, 2016).

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:

$$L_i(c) \equiv (-p_{ic}, 1 - \mu_i; -p_{ic} - c, \mu_i)\,.$$

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

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:

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

where w_i is the household's wealth and u_i is its Bernoulli utility function. In this model, aversion to risk is determined by the shape of the utility function u_i. 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 general model developed in Sections 2 and 3, 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.[9] Per Assumptions 2.1 and 4.2, we posit that $\nu_i \sim P(\boldsymbol{\gamma}(\mathbf{t}_i))$, where P is specified below in Assumption 4.3(I), and that for all $c \in \mathcal{D}$,

$$W(\mathbf{x}_{ic}, \nu_i) = \frac{(1-\mu_i)(1-\exp(\nu_i p_{ic})) + \mu_i(1-\exp(\nu_i(p_{ic}+c)))}{\nu_i} \tag{4.1}$$

and $U_i(c) = W(\mathbf{x}_{ic}, \nu_i),\ (\mathbf{x}_{ic}, \boldsymbol{\nu}_i) - a.s.$

Observe that, by equation (4.1), 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 subsamples 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 Assumptions 2.2(II) and 3.1, we assume that $\Pr(\ell_i \geqslant \kappa) = 1$ for every household $i \in \mathcal{I}$, where $\ell_i = |C_i|$ and $\kappa \geqslant 2$, and that ℓ_i conditional on $(\mathbf{x}_i, \nu_i)$ follows a discrete distribution $(\pi_q(\mathbf{t}_i), q = \kappa, \ldots, |\mathcal{D}|)$ as in equation (3.6). We could specify $\pi_q(\mathbf{t}_i) = f(\mathbf{t}_i; \boldsymbol{\eta})$ where the functional form of f is known up

[9] In terms of the general notation used in Sections 2 and 3, $\mathbf{s}_i = (\mu_i, \mathbf{t}_i)$, $\mathbf{z}_{ic} = p_{ic}$, and $\boldsymbol{\nu}_i = \nu_i$.

to $\boldsymbol{\eta} \in H$. Instead, we account for observed heterogeneity in choice set size nonparametrically by conducting the analysis separately on subsamples based on $\mathbf{t}_i$. To simplify notation, we suppress below the dependence of π_q on $\mathbf{t}_i$. Let $\boldsymbol{\pi} \equiv (\pi_q, q = \kappa, \ldots, |\mathcal{D}|)$.

We close the model with two final assumptions.

Assumption 4.3 (Heterogeneity Restrictions):

(I) Conditional on $\mathbf{t}_i$, ν_i follows a Beta distribution on $[0, 0.02]$ 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 $[0, 0.02]$. The main attraction of the Beta distribution is its flexibility. 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 0.02 is conservative both as a theoretical matter and in light of prior empirical estimates in similar settings (see, e.g., Cohen and Einav 2007; Sydnor 2010; Barseghyan et al. 2011, 2013, 2016).

Assumption 4.3(II) posits that $\Pr(\ell_i \in \{3, 4, 5\}) = 1$ for every household $i \in \mathcal{I}$. In other words, it assumes that the size of every household's choice set is either full, full-1, or full-2. In our setting the feasible set contains five alternatives. For reasons we explain in Section 4.2, we can rule out $\kappa = 5$ and $\kappa = 4$ and we set $\kappa = 3$ to balance a tradeoff between the model's empirical content and its explanatory power.

Given $(\mathbf{x}_i, \nu_i)$ and choice set $C_i = G \subseteq \mathcal{D}$, household i's optimal deductible choice is

$$d_i^*(G; \mathbf{x}_i, \nu_i) = \arg\max_{c \in G} W(\mathbf{x}_{ic}, \nu_i).$$

Given κ, the set of optimal deductible choices for all possible realizations $G \subseteq \mathcal{D}$, $|G| \geqslant \kappa$, is

$$D_\kappa^*(\mathbf{x}_i, \nu_i) = \bigcup_{G \subseteq \mathcal{D}: |G| \geqslant \kappa} \Big\{ d_i^*(G; \mathbf{x}_i, \nu_i) \Big\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \Big\{ d_i^*(G; \mathbf{x}_i, \nu_i) \Big\}. \tag{4.2}$$

The sharp identification region Θ_I of the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\pi}; \boldsymbol{\gamma}]$ is given by equation (3.7) where P is specified in Assumption 4.3(I) and D_κ^* is given by equation (4.2).

4.2 Data Description

We obtained the data from a large U.S. property and casualty insurance company. The company offers several lines of insurance, including auto. In the market where it operates, the company ranks among the top 10 writers of auto insurance. 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.

For purposes of this paper, 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.[10] This yields an initial sample of 112,011 observations. We then exclude households whose deductible choices cannot be rationalized by the model specified in Section 4.1 for any pair (ν_i, ℓ_i) such that $\nu_i \in [0, 0.02]$ and $\ell_i \in \{3, 4, 5\}$. 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 111,894 observations.

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 auto policies without actively reassessing their deductible choices. Second, we require $\nu_i \in [0, 0.02]$ for the reasons stated in Section 4.1.

Third, we require $\ell_i \in \{3, 4, 5\}$—i.e., we assume $\kappa = 3$—to balance a tradeoff between the model's empirical content (as measured by the size of Θ_I) and its explanatory power (as measured by the fraction of rationalizable households).[11] As noted previously in connection with Assumption 2.2(II), we must assume $\kappa > 1$ for the model to have any empirical content. If $\kappa = 1$ the model simply posits that a household's choice set comprises its chosen alternative and Θ_I is wholly uninformative as it comprises the entire parameter space Θ. At the same time we must assume $\kappa < 5$ because $\kappa = 5$ (even without the Beta assumption) is rejected by the data. This is because the feasible set contains a suboptimal alternative—for virtually every household at all $\nu_i \in [0, 0.02]$—that nevertheless is chosen by a sizable percentage of households; see Section 4.3. As κ decreases between 4 and 2 the model gains

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

[11] Again, our rationalizability check relies on the assumption that $\nu_i \in [0, 0.02]$ but not on the assumption that ν_i follows a Beta distribution.

explanatory power but loses empirical content. At $\kappa = 3$ the model achieves near maximum explanatory power—it can rationalize 99.9 percent of the initial sample—without losing too much empirical content: Θ_I is partially identified but still informative, as we demonstrate in Section 5. Moving down to $\kappa = 2$ would further decrease the model's empirical content with virtually no compensating gain in explanatory power. Moving up to $\kappa = 4$ would increase the model's empirical content with only a small loss in explanatory power—the model could still rationalize 99.7 percent of the initial sample. It turns out, however, that $\kappa = 4$ with the Beta assumption is rejected by the data; see Table 5.3. For these reasons, we set $\kappa = 3$.

Fourth, the company uses the same procedure to generate each household's pricing menu. The company first determines the household's base price, $\bar{p}_i$, according to a proprietary rating function. It then generates the household's pricing menu, $\mathbf{p}_i = (p_{ic}, c \in \mathcal{D})$, according to a proprietary multiplication rule, $p_{ic} = g(c)\bar{p}_i + \zeta$, where g is a decreasing positive function and ζ is a small positive scalar. The multipliers $(g(c), c \in \mathcal{D})$, known as the deductible factors, and the scalar ζ, known as the expense fee, are the same for every household. We observe each household's base price as well as the deductible factors and the expense fee.

Fifth, we construct the households' claim probabilities using the company's claims data. We begin by estimating how claim rates depend on observables. In an effort to obtain the most precise estimates, we use the full set of auto collision data, which comprises 1,349,853 household-year records. For each household-year record, the data list the number of claims filed by the household in that year. We assume that household i's 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 treat the claim rates as latent random variables and assume that $\ln \lambda_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + \varepsilon_i$, where $\mathbf{X}_{it}$ is a large vector of observables and $\exp(\varepsilon_i)$ follows a Gamma distribution with unit mean and variance ϕ. We perform Poisson panel regressions with random effects to obtain maximum likelihood estimates of $\boldsymbol{\beta}$ and ϕ.[12] Next, we use the regression results to assign claim probabilities to the households in the analysis sample. 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.[13] In principle, a household may experience one or more claims during the policy period. In the model, we assume that households disregard the possibility of experiencing more than one claim (Assumption 4.1(I)). Given this assumption, 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.

[12] The estimates are reported in Table S.1 of the Supplemental Material.

[13] 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.

Table 4.1: Summary Statistics

	Mean	Std. dev.	5th pctl.	Median	95th pctl.
Deductible choice (dollars)	439	178	200	500	500
Pricing menus:					
p_{500}	217	137	80	182	477
$p_{250} - p_{500}$	66	42	23	55	146
$p_{500} - p_{1000}$	49	32	17	41	109
Claim probability (annual)	0.088	0.030	0.047	0.084	0.142
Demographic characteristics:					
Female	0.469	0.499	0	0	1
Age (years)	48.1	16.6	24.5	45.9	76.8
Insurance score	731	114	554	725	934

Notes: Analysis sample (111,894 observations). Pricing statistics are annual amounts in nominal dollars. Demographic statistics are for the principal driver.

Tables 4.1 and 4.2 present descriptive statistics for the analysis sample. Table 4.1 summarizes the households' deductible choices, pricing menus, claim probabilities, and demographic characteristics. Table 4.2 reports the sample distribution of deductible choices for the full sample and for selected subsamples based on gender, age, and insurance score. In Table 4.2 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.

Table 4.2 also reports the sample distribution of deductible choices by quartiles of base price and claim probability. The patterns are largely as expected. Within a claim probability quartile the demand for low deductibles (\$100, \$200, and \$250) decreases, and the demand for high deductibles (\$500 and \$1000) increases, as the base price quartile increases. Within a base price quartile the demand for low deductibles increases, and the demand for high deductibles decreases, as the claim probability quartile increases. The only exception is the top base price quartile, within which the demand for low deductibles decreases, and the demand for high deductibles increases, as the claim probability quartile increases. A reasonable explanation for this anomalous pattern is that, within the top base price quartile, the rate at which base price increases with claim probability is sufficiently high that the price effect (which decreases demand for low deductibles and increases demand for high deductibles) dominates the risk effect (which increases demand for low deductibles and decreases demand for high deductibles).

Table 4.2: Deductible Choices

		Percent choosing deductible				
	Observations	\$100	\$200	\$250	\$500	\$1000
Full sample	111,894	1.1	15.2	13.7	65.4	4.6
Male	59,476	1.0	14.9	12.9	65.9	5.4
Female	52,418	1.1	15.5	14.7	64.8	3.8
Young	36,932	0.1	6.9	10.7	77.1	5.2
Old	38,049	2.5	26.2	16.7	51.0	3.6
Low insurance score	37,090	0.4	10.1	12.7	72.2	4.6
High insurance score	38,368	1.8	20.9	14.6	58.1	4.6
$\bar{p}^{Q1}, \mu^{Q1}$	13,352	2.8	29.2	18.7	46.9	2.4
$\bar{p}^{Q1}, \mu^{Q2}$	7,669	3.1	29.7	19.8	45.8	1.7
$\bar{p}^{Q1}, \mu^{Q3}$	4,721	3.0	30.4	21.7	43.8	1.0
$\bar{p}^{Q1}, \mu^{Q4}$	2,130	2.9	32.2	23.7	41.0	0.2
$\bar{p}^{Q2}, \mu^{Q1}$	7,668	0.8	17.2	15.1	62.6	4.2
$\bar{p}^{Q2}, \mu^{Q2}$	8,065	0.9	17.7	16.6	62.3	2.4
$\bar{p}^{Q2}, \mu^{Q3}$	6,952	0.9	18.3	18.2	60.8	1.9
$\bar{p}^{Q2}, \mu^{Q4}$	5,167	1.0	19.1	19.9	58.8	1.2
$\bar{p}^{Q3}, \mu^{Q1}$	4,785	0.3	9.8	11.4	73.5	5.1
$\bar{p}^{Q3}, \mu^{Q2}$	7,153	0.5	9.1	12.4	73.9	4.0
$\bar{p}^{Q3}, \mu^{Q3}$	7,939	0.4	10.2	12.7	73.2	3.5
$\bar{p}^{Q3}, \mu^{Q4}$	8,214	0.3	10.0	13.9	73.7	2.2
$\bar{p}^{Q4}, \mu^{Q1}$	2,168	0.2	4.6	5.9	80.5	8.9
$\bar{p}^{Q4}, \mu^{Q2}$	5,087	0.1	3.6	6.2	80.5	9.6
$\bar{p}^{Q4}, \mu^{Q3}$	8,361	0.0	3.3	5.0	81.9	9.8
$\bar{p}^{Q4}, \mu^{Q4}$	12,463	0.1	3.2	4.9	80.1	11.7

Notes: Analysis sample. 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. Q superscripts refer to quartiles of base price and claim probability.

4.3 Evidence of Heterogeneity in Choice Sets

A key feature of our data is that the \$200 deductible is a suboptimal alternative for virtually every household in our sample at all $\nu \in [0, 0.02]$.[14] In particular, \$200 is dominated by \$100 or \$250, depending on the household's claim probability.

To see why \$200 is a suboptimal alternative, consider a risk neutral household ($\nu = 0$) with claim probability μ. The household prefers \$200 to \$100 if and only if

$$\mu < \frac{p_{100} - p_{200}}{200 - 100} \equiv UB,$$

[14]Evaluating $W(\mathbf{x}_{ic}, \nu_i)$ in equation (4.1) for all 111,894 households over a fine grid of ν_i, we find that the \$200 deductible is optimal in 0.001 percent of cases, all of which entail $\nu_i > 0.012$. 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).

and prefers \$200 to \$250 if and only if

$$\mu > \frac{p_{200} - p_{250}}{250 - 200} \equiv LB.$$

In our data $p_{100} - p_{200} = p_{200} - p_{250}$ for all households. For the risk neutral household, therefore, $UB < LB$, which implies that 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, and indeed for virtually every household in our sample \$200 is suboptimal at all $\nu \in [0, 0.02]$. This logic applies whether risk aversion is driven by diminishing marginal utility as in expected utility theory or by probability weighting as in rank-dependent expected utility theory.

Yet 15.2 percent of households in our sample choose the \$200 deductible. At the same time, only 1.1 percent choose \$100 and 13.7 percent choose \$250. Hence, the demand for \$100 and \$250, separately and together, are less than the demand for \$200. This pattern is even more pronounced within certain subsamples, including households with old principal drivers, households with high insurance scores, households with base prices in the first quartile, and households with claim probabilities in the first quartile and base prices in the first or second quartiles; see Table 4.2.

Heterogeneity in 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.

By contrast, we now establish that many standard discrete choice models, such as the mixed logit model (e.g., McFadden and Train 2000; Train 2009) and the trembling hand model (e.g., Harless and Camerer 1994; Wilcox 2008), cannot explain the choice probabilities in our data. Consider the following mixed logit (MixL) and trembling hand (TH) models. Note that *none* of the results in the remainder of this section rely on the assumptions of the empirical model set forth in Section 4.1, including the assumption that households have expected utility preferences with Beta distributed, constant absolute risk aversion.

Definition 4.1 (MixL): *$U_i(c) = W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) + \epsilon_{ic}$, where ϵ_{ic} is a random i.i.d. disturbance that follows a Type 1 Extreme Value distribution and is independent of $(\mathbf{x}_{ic}, \boldsymbol{\nu}_i)$, and $\kappa = |\mathcal{D}|$.*

Definition 4.2 (TH): *$U_i(c) = W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, $\kappa = 1$, $C_i \perp \boldsymbol{\nu}_i$, $\Pr(C_i = \mathcal{D}) = 1 - \varpi$, and $\Pr(C_i = \{c\}) = \frac{\varpi}{|\mathcal{D}|}$ for all $c \in \mathcal{D}$.*[15]

Neither MixL nor TH can explain $\Pr(d = 100|\mathbf{x}) + \Pr(d = 250|\mathbf{x}) < \Pr(d = 200|\mathbf{x})$.

Claim 4.1: *Take the model in Section 2. Suppose that for a given $c \in \mathcal{D}$ there exist $a, b \in \mathcal{D}$, $a \neq b \neq c$, such that, $\boldsymbol{\nu} - a.s.$, $W(\mathbf{x}_a, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})$ or $W(\mathbf{x}_b, \nu; \boldsymbol{\delta}) > W(\mathbf{x}_c, \nu; \boldsymbol{\delta})$. Then for any distribution of $\boldsymbol{\nu}$ with support $\mathcal{V}$:*

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

(II) Under TH, $\min\{\Pr(d = a|\mathbf{x}), \Pr(d = b|\mathbf{x})\} \geqslant \Pr(d = c|\mathbf{x}),\ \mathbf{x} - a.s.$

The intuition behind Claim 4.1(II) is straightforward. Under TH, alternative c is chosen only as a result of a tremble ($C_i = \{c\}$) whereas alternative a or b may be chosen as a result of a tremble or because it is optimal ($C_i = \mathcal{D}$). Because all trembles are equiprobable, the choice probabilities of a and b can never be less than the choice probability of c.

Claim 4.1(I) follows from the fact that MixL 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 4.1 (Conditional Rank Order Property): *For all $c, c' \in \mathcal{D}$, $\Pr(d_i = c'|\mathbf{x}_i, \boldsymbol{\nu}_i) \geqslant \Pr(d_i = c|\mathbf{x}_i, \boldsymbol{\nu}_i)$ if and only if $W(\mathbf{x}_{ic'}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \geqslant W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, $\boldsymbol{\nu}_i - a.s.$*

To see that MixL satisfies Property 4.1, note that under MixL,

$$\Pr(d_i = c'|\mathbf{x}_i, \boldsymbol{\nu}_i) = \frac{\exp(W(\mathbf{x}_{ic'}, \boldsymbol{\nu}_i; \boldsymbol{\delta}))}{\sum_{c \in \mathcal{D}} \exp(W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}))}. \tag{4.3}$$

Property 4.1 follows from equation (4.3) for any distribution of $\boldsymbol{\nu}_i$ with support $\mathcal{V}$, and Claim 4.1(I) follows from Property 4.1 by integrating with respect to the distribution of $\boldsymbol{\nu}_i$. Indeed, any discrete choice model that satisfies Property 4.1—including, inter alia, the MixL model, the conditional logit model (McFadden 1974), the semiparametric random utility model of Manski (1975), and the multinomial probit model (e.g., Hausman and Wise 1978)—is incapable of explaining the choice probabilities in our data.

Claim 4.1(II) highlights the fact that not all forms of choice set heterogeneity can explain the choice patterns in our data. To further illustrate this point, consider variants of the two models of choice set formation referenced at the end of Section 2. The first, which we call the

[15] We could condition the choice set distribution on $\mathbf{t}_i$ and the result in Claim 4.1(II) would go through mutatis mutandis.

Uniform Random (UR) model, posits that a household's choice set C_i is drawn uniformly at random from the feasible set $\mathcal{D}$, conditional on $\ell_i = q$ for $q \geqslant \kappa$ (cf. Dardanoni et al. 2018). The second, which we call the Alternative Specific Random (ASR) model, posits that each alternative $c \in \mathcal{D}$ enters a household's choice set C_i with probability $\varphi(c)$ independently of other alternatives, conditional on $\ell_i \geqslant \kappa$ (cf. Manski 1977; Manzini and Mariotti 2014).

Definition 4.3 (UR): $\Pr(C_i = G|\ell_i = q) = \binom{|\mathcal{D}|}{q}^{-1}$ *for all* $G \subseteq \mathcal{D}$, $|G| = q$, $q \geqslant \kappa$; *and* $C_i \perp \boldsymbol{\nu}_i$.

Definition 4.4 (ASR): $\Pr(C_i = G|\ell_i \geqslant \kappa) = \Pr(C_i = G)/(1 - \sum_{G\subseteq\mathcal{D}:|G|<\kappa} \Pr(C_i = G))$ *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* $C_i \perp \boldsymbol{\nu}_i$.[16]

Our model coupled with ASR can explain $\Pr(d = 100|\mathbf{x}) + \Pr(d = 250|\mathbf{x}) < \Pr(d = 200|\mathbf{x})$, but our model coupled with UR cannot.

Claim 4.2: *Take the model in Section 2. Suppose that for a given* $c \in \mathcal{D}$ *there exist* $a, b \in \mathcal{D}$, $a \neq b \neq c$, *such that,* $\boldsymbol{\nu} - a.s.$, $W(\mathbf{x}_a, \boldsymbol{\nu}; \boldsymbol{\delta}) > W(\mathbf{x}_c, \boldsymbol{\nu}; \boldsymbol{\delta})$ *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) Under UR, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) > \Pr(d = c|\mathbf{x})$, $\mathbf{x} - a.s.$

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

Claim 4.2(I) follows from the fact that our model coupled with UR satisfies Property 4.1. It is easy to see why. 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 ℓ_i, are equiprobable, c' is chosen more frequently than c.

We can establish Claim 4.2(II) with a trivial example. Suppose $\varphi(a) = \varphi(b) = 0$ and $\varphi(c) = 1$. Then $\Pr(d = a|\mathbf{x}_i) = \Pr(d = b|\mathbf{x}_i) = 0$ and $\Pr(d = c|\mathbf{x}_i) > 0$ provided that there exists a positive measure of values of $\boldsymbol{\nu}_i \in \mathcal{V}$ such that $W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}) > W(\mathbf{x}_{ic'}, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ for all $c' \in \mathcal{D}\setminus\{a, b\}$, $c' \neq c$. More generally, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) < \Pr(d = c|\mathbf{x})$ is possible as long as $\varphi(a)$ and $\varphi(b)$ are sufficiently low, $\varphi(c)$ is sufficiently high, and c is the best alternative in $\mathcal{D}\setminus\{a, b\}$ for some positive measure of values of $\boldsymbol{\nu}_i \in \mathcal{V}$.

Remark 4.1: There is another noteworthy distinction between our model and models of the form $U_i(c) = W_i(c) + \epsilon_{ic}$ where $W_i(c)$ is nonrandom and represents expected utility

[16] We could condition the choice set distribution on $\mathbf{t}_i$ and the result in Claim 4.2(II) would go through mutatis mutandis.

preferences with CARA and ϵ_{ic} is a random i.i.d. disturbance that follows a continuous and strictly increasing distribution. Models in this class violate a basic monotonicity property: given C_i, as risk aversion increases the choice probabilities of the riskier alternatives decrease at first but eventually increase (Apesteguia and Ballester 2018). This happens because differences in expected utilities converge to zero as risk aversion increases, allowing differences in disturbances to determine choices. The problem is that models in this class are cardinal in W. Our model avoids this pitfall because it is ordinal in W.

5 Empirical Findings

Our empirical application is motivated in part by the suboptimal choices and related choice patterns discussed above, which are suggestive of unobserved heterogeneity in choice sets. At the same time, it is also motivated by a persistent finding in prior empirical studies of risk preferences which assume full sized choice sets. These studies, many of which estimate expected utility models and some of which estimate non-expected utility models, tend to find that average risk aversion is quite high—arguably implausibly high—and that heterogeneity in risk aversion is rather large as well. Two recent examples that utilize data on deductible choices in property insurance are Cohen and Einav (2007) and Barseghyan et al. (2013).[17] Our empirical application is motivated in large part by the hypothesis that the assumption of full sized choice sets may be driving this finding and that allowing for heterogeneity in choice sets may yield more credible estimates of risk preferences.

We begin with a brief explanation of our empirical methods (Section 5.1). We then apply Theorem 3.1 to learn about the distribution of risk preferences (Section 5.2). After that, we apply Corollary 3.1 to learn about the distribution of choice set size (Section 5.3) and Theorem 3.2 to assess the welfare cost of limited choice sets (Section 5.4).

5.1 Summary of Empirical Methods

The inequalities in equations (3.5) and (3.7) and in the constraints in problem (3.8) need to hold $(\mu_i, \mathbf{t}_i, \mathbf{p}_i) - a.s.$ As indicated in Section 4.1, we account for observed heterogeneity by conducting our analysis separately for subsamples based on $\mathbf{t}_i$. In keeping with the common practice in the empirical literature on partial identification (e.g., Ciliberto and Tamer 2009), we aggregate the inequalities within each equation by discretizing the support of $(\mu_i, \mathbf{p}_i)$ in

[17] For a review of these and other studies in the literature on estimating risk preferences using field data, see Barseghyan et al. (2018).

bins B_j, $j = 1, \dots, J$. We estimate the left hand side of the aggregated inequalities by

$$\widehat{\Pr}(d_i \in K | (\mu_i, \mathbf{p}_i) \in B_j) = \frac{\sum_{i=1}^n \mathbf{1}(d_i \in K, (\mu_i, \mathbf{p}_i) \in B_j)}{\sum_{i=1}^n \mathbf{1}((\mu_i, \mathbf{p}_i) \in B_j)}.$$

The right hand side is a model defined function of $\boldsymbol{\theta}$. For reasons we explain below, we use 64 bins: 8 quantiles each for μ_i and $\bar{p}_i$.[18] Where possible, we leverage Appendix A, Theorem A.2 and other strategies to eliminate redundant inequalities and reduce the number of inequalities that need to be checked.

We use the generalized moment selection procedure introduced by Andrews and Soares (2010) [hereafter, AS] to obtain confidence sets that asymptotically uniformly cover $\boldsymbol{\theta} \in \Theta_I$ with probability 95 percent. Under Theorems 3.1 and 3.2, $\boldsymbol{\theta} = (\gamma_1, \gamma_2)$. Under Corollary 3.1, $\boldsymbol{\theta} = (\gamma_1, \gamma_2, \pi_3, \pi_4, \pi_5)$. 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 smooth functions of γ_1 and γ_2 (e.g., $\mathrm{E}(\nu_i) = (0.02 \times \gamma_1)/(\gamma_1 + \gamma_2)$) and for π_3, π_4, and π_5. The only exception is that we report 95 percent confidence intervals for percentiles of ν_i based on projections of the AS confidence set. In all cases we use 1,000 bootstrap replications. We review the AS and KMS methods in Appendix B.

Although it is common practice in the empirical literature on partial identification to discretize the support of the covariates, there is no established best practice for how to do so. We proceed as follows. We construct the AS 95 percent confidence set for $\boldsymbol{\theta} = (\gamma_1, \gamma_2)$, using the entire sample, when the support of $(\mu_i, \mathbf{p}_i)$ is discretized in 16, 25, 64, and 100 bins; see Figure 5.1. As we refine the discretized suppport, the resulting inequalities are different aggregations of the inequalities in equations (3.5), (3.7), and (3.8) which hold $(\mu_i, \mathbf{p}_i) - a.s.$ Nevertheless it is natural to expect that more information is harnessed with each refinement, and that is what we observe as we move from 16 to 64 bins or from 25 to 100 bins (i.e., as we double the number of quantiles for μ_i and $\bar{p}_i$ starting from quartiles or quintiles). With limited data, however, there is statistical uncertainty which increases with each refinement and tends to enlarge the confidence set. We find that the confidence sets resulting from 64 and 100 bins are essentially the same. Because the former entails fewer inequalities, we use the 64 bins based on 8 quantiles each for μ_i and $\bar{p}_i$ throughout our empirical analysis.

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) and the constraints in problem (3.8) are satisfied. This implies that we fail to reject the hypothesis that our empirical model is

[18] Given how the company generates $\mathbf{p}_i$, a household's base price $\bar{p}_i$ is a sufficient statistic for $\mathbf{p}_i$.

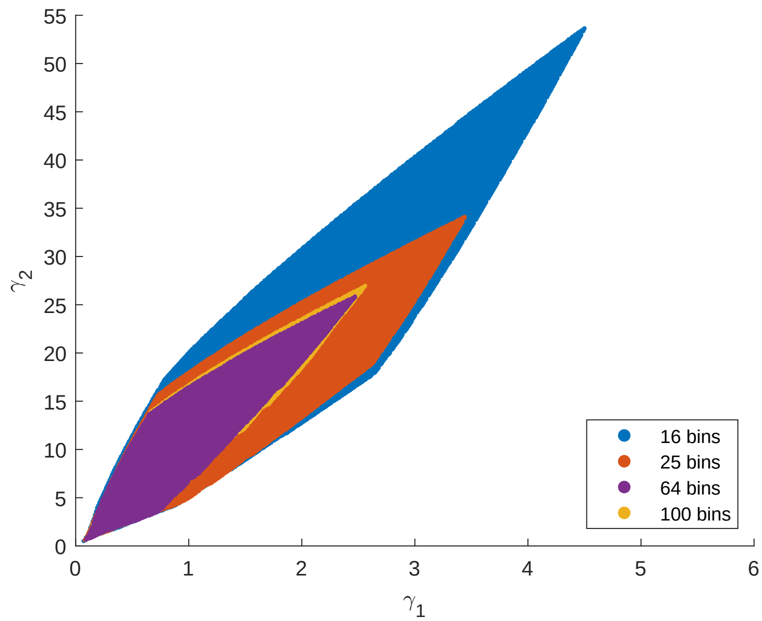


Figure 5.1: Confidence set for $\boldsymbol{\theta}$.

Note: The figure depicts the AS 95 percent confidence set for $\boldsymbol{\theta} = (\gamma_1, \gamma_2)$ when the support of $(\mu_i, \mathbf{p}_i)$ is discretized in 16, 25, 64, and 100 bins.

correctly specified (though, of course, this does not guarantee that it is correctly specified). For methods to test for misspecification in moment inequality models, see Bugni et al. (2015).

5.2 Risk Preferences

Columns (1) and (2) of Table 5.1 present KMS 95 percent confidence intervals for $\mathrm{E}(\nu_i)$ and $\mathrm{Var}(\nu_i)$, respectively, for the full sample and for selected subsamples based on gender, age, and insurance score. Recall that ν_i is the coefficient of absolute risk aversion. Focusing on the lower bounds, we find that for the full sample the mean of absolute risk aversion is as low as 0.00091, with a variance as low as 0.0000012. When we split the sample by gender, we find only small differences: the lower bounds are 6 to 10 percent higher for households with female principal drivers than for households with male principal drivers. When we split the sample by age and insurance score, by contrast, we find large differences: the lower bounds are 2.8 to 4.6 times higher for households with old principal drivers than for households with young principal drivers, and 2.4 to 5.9 times higher for households with high insurance scores than for households with low insurance scores.

Table 5.1: Risk Preferences and Welfare

	(1) E(ν)		(2) Var(ν)		(3) Welfare cost
	LB	UB	LB	UB	UB
Full sample	$9.1 \cdot 10^{-4}$	$3.4 \cdot 10^{-3}$	$1.2 \cdot 10^{-6}$	$2.5 \cdot 10^{-5}$	\$54.49
Male	$8.9 \cdot 10^{-4}$	$3.1 \cdot 10^{-3}$	$1.0 \cdot 10^{-6}$	$2.4 \cdot 10^{-5}$	\$54.24
Female	$9.4 \cdot 10^{-4}$	$3.7 \cdot 10^{-3}$	$1.1 \cdot 10^{-6}$	$2.6 \cdot 10^{-5}$	\$54.61
Young	$4.0 \cdot 10^{-4}$	$3.0 \cdot 10^{-3}$	$2.2 \cdot 10^{-7}$	$2.2 \cdot 10^{-5}$	\$76.20
Old	$1.1 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$	$1.0 \cdot 10^{-6}$	$3.4 \cdot 10^{-5}$	\$36.25
Low insurance score	$4.0 \cdot 10^{-4}$	$3.3 \cdot 10^{-3}$	$1.6 \cdot 10^{-7}$	$2.2 \cdot 10^{-5}$	\$59.68
High insurance score	$9.7 \cdot 10^{-4}$	$4.9 \cdot 10^{-3}$	$9.5 \cdot 10^{-7}$	$3.3 \cdot 10^{-5}$	\$47.12

Notes: KMS 95 percent confidence intervals. LB and UB denote lower bound and upper bound, respectively. The lower bound on the welfare cost is zero by definition.

Table 5.2: Interpretation of E(ν) and Var(ν)

	E(ν)	Risk premium	25th pctl.	75th pctl.
This paper:				
Lower bound of CI	$9.1 \cdot 10^{-4}$	\$ 52	$3.6 \cdot 10^{-9}$	$1.0 \cdot 10^{-3}$
Upper bound of CI	$3.4 \cdot 10^{-3}$	\$300	$9.7 \cdot 10^{-4}$	$5.1 \cdot 10^{-3}$
MixL	$1.7 \cdot 10^{-3}$	\$122	$1.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
TH	$1.7 \cdot 10^{-3}$	\$113	$9.7 \cdot 10^{-4}$	$2.2 \cdot 10^{-3}$
UR	$1.7 \cdot 10^{-3}$	\$116	$1.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
ASR	$2.6 \cdot 10^{-3}$	\$212	$7.0 \cdot 10^{-4}$	$3.8 \cdot 10^{-3}$
Cohen and Einav (2007):				
Benchmark model	$6.7 \cdot 10^{-3}$	\$558	$2.3 \cdot 10^{-6}$	$2.9 \cdot 10^{-4}$
CARA model	$3.1 \cdot 10^{-3}$	\$267	NR	NR
Barseghyan et al. (2013):				
Model 4	$1.5 \cdot 10^{-3}$	\$ 97	$7.2 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$
CARA model	$1.1 \cdot 10^{-3}$	\$ 68	NR	NR

Notes: Confidence intervals for E(ν) and risk premium are KMS 95 percent confidence intervals. Confidence intervals for 25th and 75th percentiles of ν are based on projections from the AS 95 percent confidence set for $\boldsymbol{\theta}$. Risk premium is calculated for an agent with CARA expected utility preferences and a lottery that yields a loss of \$1000 with probability 10 percent. CI = confidence interval. NR = not reported.

To help interpret and provide context for the KMS 95 percent confidence intervals for $\mathrm{E}(\nu_i)$ and $\mathrm{Var}(\nu_i)$, Table 5.2 reports: (i) point estimates of $\mathrm{E}(\nu_i)$ obtained under eight comparator models; (ii) the risk premium, for an agent with CARA expected utility preferences, of a lottery that yields a loss of \$1000 with probability 10 percent (and no gain or loss with probability 90 percent), computed at the lower and upper bounds of the KMS 95 percent confidence interval for $\mathrm{E}(\nu_i)$ and at each of the comparison point estimates of $\mathrm{E}(\nu_i)$; and (iii) 95 percent confidence intervals for the 25th and 75th percentiles of ν_i based on projections of the AS 95 percent confidence set for $\boldsymbol{\theta}$, as well as point estimates of the 25th and 75th percentiles of ν_i obtained under six of the eight comparator models.[19]

The eight comparator models are the MixL, TH, UR, and ASR models described in Section 4.3, two models in Cohen and Einav (2007) (their benchmark and CARA models), and two models in Barseghyan et al. (2013) (their Model 4 and CARA model). Cohen and Einav (2007) estimate the distribution of absolute risk aversion in a parametric expected utility model with observed and unobserved heterogeneity in risk preferences using data on deductible choices in Israeli auto insurance. The Bernoulli utility function is a second-order Taylor expansion in their benchmark model and a CARA utility function in their CARA model. Barseghyan et al. (2013) estimate the distribution of absolute risk aversion and probability distortions in a parametric rank-dependent expected utility model with heterogeneity in risk preferences using data on deductible choices in U.S. auto and home insurance. Their data and our data are sourced from the same company. The Bernoulli utility function is a second-order Taylor expansion in their Model 4 and a CARA utility function in their CARA model. They allow for observed heterogeneity in their CARA model and for observed and unobserved heterogeneity in their Model 4. The estimates reported in Table 5.2 for the models in Cohen and Einav (2007) and Barseghyan et al. (2013) are the estimates they report based on their data. We estimate the MixL, TH, UR, and ASR models using our data.

The main takeaway from Table 5.2 is that the lower bounds of the confidence intervals for $\mathrm{E}(\nu_i)$ and the 25th percentile of ν_i are substantially smaller than the corresponding estimates obtained under all of the comparator models that model risk preferences by expected utility theory (i.e., all but the two models in Barseghyan et al. (2013)). The lower bound of the confidence interval for the 75th percentile of ν_i is also less than the corresponding estimates obtained under all but one of these comparator models. However, because the distribution of ν_i is right skewed, the 25th percentile is the more relevant point of comparison. Even the upper bound of the confidence interval for the 25th percentile of ν_i is less than the corresponding estimates obtained under all but two of these comparator models, though not

[19]Neither Cohen and Einav (2007) nor Barseghyan et al. (2013) report these percentiles for their CARA models.

surprisingly the upper bound for $\mathrm{E}(\nu_i)$ is greater than the corresponding estimates obtained under all but one of these comparator models. Altogether, the confidence intervals suggest that, if one properly allows for heterogeneity in choice sets, the data can be explained by expected utility theory with lower and more homogeneous 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 from Table 5.2 comes from the comparison of the confidence intervals for $\mathrm{E}(\nu_i)$ and the 25th and 75th percentiles of ν_i to the corresponding point estimates reported by Barseghyan et al. (2013), who model risk preferences by rank-dependent expected utility theory. The lower bounds of the confidence intervals are all considerably less than the corresponding point estimates reported by Barseghyan et al. (2013), suggesting that heterogeneity in choice sets may be a viable substitute or complement to heterogeneity in probability distortions in terms of explaining risky choices. Indeed, one can view both types of heterogeneity as forms of heterogeneity in inattention—inattention to alternatives and inattention to probabilities (Gabaix 2018).

Figure 5.2 depicts a 95 percent confidence set for an outer region of admissible probability density functions of ν_i based on the AS confidence set for $\boldsymbol{\theta}$. It also superimposes the predicted density functions of ν_i based on maximum likelihood estimates of $\boldsymbol{\theta}$ obtained under the MixL, TH, UR, and ASR models. 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.[20]

Figure 5.2 shows that the MixL, TH, and UR models predict density functions that do not lie entirely inside the confidence set for the outer region of admissible probability density functions of ν_i based on the AS confidence set for $\boldsymbol{\theta}$. Indeed, the maximum likelihood estimates of $\boldsymbol{\theta}$ under these models are rejected by the AS test of the hypothesis that $\boldsymbol{\theta}_0 \in \Theta_I$.[21] The ASR model, by contrast, is not rejected and, in fact, its predicted density function for ν_i lies entirely inside the confidence set for the outer region of admissible probability density

[20] 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. Figure 5.2 presents the outer region of admissible probability density functions of ν_i for the full sample. Figure S.1 in the Supplemental Material presents the outer region for selected subsamples based on gender, age, and insurance score.

[21] Given our large sample size, confidence sets on maximum likelihood estimates of MixL, TH, and UR are very tight, and all values in them would be rejected if tested as $\boldsymbol{\theta}_0 \in \Theta_I$.

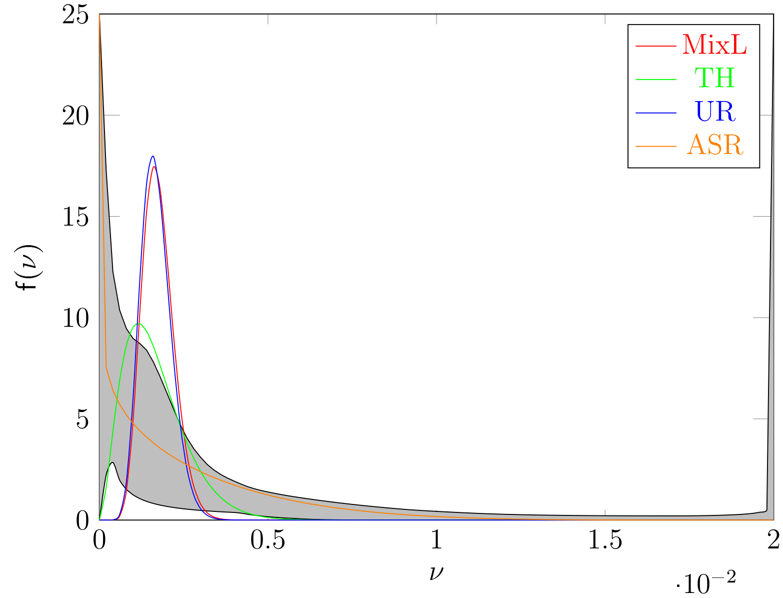


Figure 5.2: Confidence set for outer region of admissible probability density functions of ν.

Note: The figure depicts a 95 percent confidence set for an outer region of admissible probability density functions of ν_i based on the AS 95 percent confidence set for $\boldsymbol{\theta}$. It also superimposes the implied probability density functions of ν_i based on maximum likelihood estimates of $\boldsymbol{\theta}$ obtained under the MixL, TH, UR, and ASR models.

functions (though we note that this does not imply that the estimated choice set distribution obtained under the ASR model is not rejected).

5.3 Choice Set Size

Table 5.3 reports the KMS 95 percent confidence intervals for π_5, π_4, and π_3 for the full sample and the usual subsamples. The interesting quantities are the upper bounds on π_5 and π_4.[22] The former is the maximum fraction of households whose deductible choices can be rationalized with full 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 on π_3) is the minimum fraction of households who require full-2 choice sets.

[22]By construction, due to the assumption that $\kappa = 3$ (Assumption 4.3(II)), 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.

Table 5.3: Distribution of Choice Set Size

	π_5 (full)		π_4 (full-1)		π_3 (full-2)	
	LB	UB	LB	UB	LB	UB
Full sample	0.00	0.20	0.00	0.93	0.07	1.00
Male	0.00	0.20	0.00	0.93	0.07	1.00
Female	0.00	0.30	0.00	0.97	0.03	1.00
Young	0.00	0.14	0.00	0.99	0.01	1.00
Old	0.00	0.32	0.00	0.96	0.04	1.00
Low insurance score	0.00	0.28	0.00	1.00	0.00	1.00
High insurance score	0.00	0.31	0.00	0.99	0.01	1.00

Notes: KMS 95 percent confidence intervals. LB and UB denote lower bound and upper bound, respectively. By construction, because $\kappa = 3$, the lower bounds on π_5 and π_4 are zero, the lower bound for π_3 is one minus the upper bound on π_4, and the upper bound on π_3 is one.

The main result is that a large majority of households require limited choice sets (full-1 or full-2) to explain their deductible choices. For the full sample, we find that at least 80 percent of households require limited choice sets, including at least 7 percent who require full-2 choice sets. In addition, we find that: (i) more households with male principal drivers than with female principal drivers require limited choice sets (80 percent versus 70 percent), including full-2 choice sets (7 percent versus 3 percent); (ii) more households with young principal drivers than with old principal drivers require limited choice sets (86 percent versus 68 percent), though more of the latter require full-2 choice sets (1 percent versus 4 percent); and (iii) more households with low insurance scores than with high insurance scores require limited choice sets (72 percent versus 69 percent), though a bit more of the latter require full-2 choice sets (0 percent versus 1 percent).[23]

5.4 Welfare

Lastly, Column (3) of Table 5.1 reports the upper bound of the KMS 95 percent confidence interval for the solution to problem (3.8). As explained in Section 3.3, this provides a measure of the welfare cost of limited choice sets in our context. In the full sample, we find that the upper bound on this welfare cost is \$54. To put this in context, recall that the mean price

[23]When we split the full sample to form subsamples based on gender, age or insurance score, the full sample and the subsamples all have different confidence sets for $\boldsymbol{\theta}$ and, moreover, the subsamples all contain fewer observations than the full sample. Consequently, it is possible that the upper bound on π_5 for the full sample is not a weighted average of the upper bounds on π_5 for the subsamples. The same is true for the upper bound on π_4 (and, therefore, for the lower bound on π_3).

of coverage with a \$500 deductible (the modal choice) is \$217 (see Table 4.1). Thus, we find that the welfare cost of limited choice sets may be as high as 25 percent of what the average household spends on coverage. Moreover, while we do not find any meaningful difference based on gender, we find that this welfare cost may be somewhat higher/lower for households with low/high insurance scores (\$60/\$47) and considerably higher/lower for households with young/old principal drivers (\$76/\$36).

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.[24] In practice choice sets are often unobserved. Manski (1977), among others, highlights this issue.[25] In an influential paper he 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 (Manski 1977, p. 239).

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).[26] 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

[24]See McFadden (1974, p. 107): "Observed data are assumed to be generated by the *trial* of drawing an individual randomly from the population and recording his attributes, the set of alternatives available to him, and his actual choice. A sample is obtained by a sequence of independent trials...."

[25]See Manski (1977, p. 239): "Current methods for estimating the parameters of random utility functions require ex post observation of a sequence of choice problems for each of which the decision maker, choice set and chosen alternative are known. Often, however, the survey instrument used in estimation supplies the identities of the decision makers and his chosen alternative but not those of his feasible inferior alternatives." See also, e.g., Ben-Akiva (1973, pp. 83-84): "The question that remains is, therefore, how to determine the set of alternatives...that the consumer is choosing from....[I]t is likely that the actual choice is made out of only a subset of the [feasible] set. The problem is to determine this subset."

[26]For an exception, see, e.g., Horowitz and Louviere (1995).

set formation process, including with respect to the dependence or lack thereof between preferences and choices sets (conditional on observables).[27]

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.[28] More specifically, 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. The discrete choice literature is vast, spanning a diverse array of fields and subfields such as econometrics, experimental economics, microeconomics, behavioral economics, decision theory, macroeconomics, financial economics, education, labor economics, industrial organization, marketing, and transportation 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.[29] 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). More often than not, this approach is taken implicitly without discussion or justification.

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

[27] Cf. Ben-Akiva (1973, pp. 84-85): "Any determination of [the choice set] involves an a priori arbitrary criterion....Actually, every existing model explicitly or implicitly makes some a priori assumption that determines the relevant subset of alternatives."

[28] Many important papers in the theory literature—including papers on revealed preference analysis under limited attention, limited consideration, rational inattention, 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. 2019). 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. (2018), which assumes that only a single cross-section of aggregate choice shares is observed.

[29] 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."

sales, use survey data on brand awareness;[30] De los Santos et al. (2012), who study online book purchases, use survey data on web browsing;[31] 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.[32]

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;[33] 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.[34]

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 (2018), Barseghyan et al. (2019), Crawford et al. (2019), Lu (2018), and Cattaneo et al. (2019).[35]

Abaluck and Adams (2018) 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.

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

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

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

[33] 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.

[34] Heiss et al. (2016) similarly assume that a Medicare Part D insured's decision to consider alternatives to her existing prescription 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.

[35] Dardanoni et al. (2018) also take this approach. However, they rule out unobserved preference heterogeneity and focus on point identification of the choice set formation model.

Barseghyan et al. (2019) study random preference models (as opposed to classic random utility models with additive i.i.d. disturbances) which satisfy the Spence-Mirrlees single crossing property. They show that such models are point identified when coupled with variants of the ASR and UR models described above, assuming that choice sets and preferences are independent conditional on observables and that there exists an agent specific attribute with large support that shifts preferences over alternatives but does not affect choice sets.

Crawford et al. (2019) show that with panel data (or group-homogeneous cross-section data) and preferences in the logit family, point identification of preferences is possible, without 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 (2018) 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 (i.e., the monotonicity condition that $c \in^* G'$ implies $c \in^* G$ for any $G \subset G'$ such that $c \in G$), yields moment inequalities on the choice probabilities, which he uses to obtain outer regions on the model's preference parameters. He also shows that additional large support conditions, monotonicity restrictions on model implied choice probabilities, and further assumptions on the joint distribution of agents' unobserved choice sets and their observed upper and lower bounds can be used to obtain point identification.

Cattaneo et al. (2019) 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 agents' 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 2018; Barseghyan et al. 2019; Crawford et al. 2019; Cattaneo et al. 2019). Nor do we impose other restrictions on how agents' choice sets are formed (Abaluck and Adams 2018; Barseghyan et al. 2019) or evolve over time (Crawford et al. 2019), rely on exclusion restrictions or large support assumptions (Abaluck and Adams 2018; Barseghyan et al. 2019), require that the econometrician knows the composition of the smallest possible choice set for each agent (Abaluck and Adams 2018; Lu 2018), or assume that choice sets satisfy a monotonicity or other regularity condition (Lu 2018; Cattaneo et al. 2019).

Due to the parsimony of our approach we obtain partial and not point identification of the underlying model of preferences. Nevertheless, as we demonstrate in our empirical application, 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 agents' 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) and to conduct welfare analysis (without any additional assumptions). We also show how it can be used to construct tests for rejecting hypothesized models of choice set formation (given the underlying model of preferences).

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 and more homogeneous 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.

Appendices

A Theory

A.1 Additive Error Random Utility Models

The classic random utility models in the tradition of McFadden (1974), which have the form $U_i(c) = W_i(c) + \epsilon_{ic}$ where ϵ_{ic} is an additive disturbance that is agent and alternative specific, can be subsumed within our framework as follows. Let $|\boldsymbol{\nu}_i| \geqslant |\mathcal{D}| + 1$, let $\tilde{\boldsymbol{\nu}}_i$ denote the first $|\mathcal{D}|$ components of $\boldsymbol{\nu}_i$, and let $\{\mathbf{e}_c : c \in \mathcal{D}\}$ be a collection of $|\mathcal{D}| \times 1$ standard basis vectors whose cth component equals one. Then $\epsilon_{ic} = \mathbf{e}_c^\top \tilde{\boldsymbol{\nu}}_i$. For reasons we explain in Section 4.3, we dispense with ϵ_{ic} in our empirical model and focus on unobserved heterogeneity in choice sets, which we conceptualize as agent specific. However, one may conceptualize unobserved heterogeneity in choice sets as agent and alternative specific. In a classic random utility model, one may let $\epsilon_{ic} \in \{-\infty, 0\}$ for each alternative $c \in \mathcal{D}$ and allow ϵ_{ic} to be correlated with $\epsilon_{ic'}$ for any two alternatives $c, c' \in \mathcal{D}$. One would then posit that: if $\kappa = |\mathcal{D}|$ then $\epsilon_{ic} = 0$ for each alternative $c \in \mathcal{D}$; if $\kappa = |\mathcal{D}| - 1$ then $\epsilon_{ic} = -\infty$ for at most one alternative in $\mathcal{D}$ (the identity of which is left unspecified); if $\kappa = |\mathcal{D}| - 2$ then $\epsilon_{ic} = -\infty$ for at most two alternatives in $\mathcal{D}$ (the identities of which are left unspecified); and so on. This model yields that alternative c is not chosen if $\epsilon_{ic} = -\infty$, which is analogous to alternative c not being chosen when it is not contained in the agent's choice set.

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*

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

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. □

A.3 Proof of Theorem 3.1

Let $d^*_i(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ denote the model implied optimal choice for agent i with choice set G. Recall that by Assumption 2.2(II), $\Pr(C_i = G|\mathbf{x}_i, \boldsymbol{\nu}_i) = 0$ 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 $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ for C_i such that $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) \geqslant 0$ for all $G \subseteq \mathcal{D}$, $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = 0$ if $|G| < \kappa$, $\sum_{G \subseteq \mathcal{D}} \mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i) = 1$, and for all $c \in \mathcal{D}$

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

This is because for such values of $\boldsymbol{\theta}$ one can complete the model with a distribution $\mathsf{F}(\cdot; \mathbf{x}_i, \boldsymbol{\nu}_i)$ 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 $(d_i, \mathbf{x}_i)$ is a selection of the random closed set $(D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}), \mathbf{x}_i)$ if and only if the condition in equation (3.5) holds. Take $\boldsymbol{\theta}$ such that there exists a distribution $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i)$ under which equation (A.2) holds. By definition $d^*_i(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is a selection of $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$, and by Molchanov and Molinari (2018, Theorem 2.33) equation (3.5) holds. Conversely, take a value of $\boldsymbol{\theta}$ for which the inequalities in equation (3.5) are satisfied. Then, by Theorem A.1, there exists a selection $\tilde{d}_i(G)$ of $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ such that $\Pr(d_i = c|\mathbf{x}_i) = \Pr(\tilde{d}_i(G) = c|\mathbf{x}_i)$ for some G such that $|G| \geqslant \kappa$. Let $\mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\nu}_i)$ equal 1 for one such set G such that $\tilde{d}_i(G) = c$, and equal 0 for all other $G \subseteq \mathcal{D}$. Then equation (A.2) holds. □

A.4 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 non-empty proper subsets of $\mathcal{D}$, to a smaller collection. The reduced collection that suffices for Theorem 3.1 also suffices for Theorem 3.2.

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 $\mathbb{K} = \{K : K \subsetneq \mathcal{D}\}$. Then:*

1. *For any set $K \in \mathbb{K}$ such that $|K| \geqslant \kappa$, set $\mathbb{K} = \mathbb{K} \backslash K$;*

2.(1) *For any 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 $\mathbb{K} = \mathbb{K} \backslash K$;*[36]

2.(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 2.1 and 2.(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. Recall that the set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ comprises the $|\mathcal{D}| - \kappa + 1$ best alternatives in $\mathcal{D}$. Step 1 then follows because any set $K : |K| \geqslant \kappa$ includes at least the $(|\mathcal{D}| - \kappa + 1)$-th best alternative for all realizations of $\boldsymbol{\nu}$ in $\mathcal{V}$, so that $\Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing) = 1$ and the inequality in equation (3.5) holds mechanically. Step 2.(1) follows because under the stated condition, again $\Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing) = 1$. Step 2.(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}. □

[36] 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 on.

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.1: *Let Assumptions 2.1 and 2.2 hold. Suppose all possible realizations of* $D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta})$ *are given by adjacent elements of* $\mathcal{D}$*, as* $\{c_j, c_{j+1}, \ldots, c_{j+|\mathcal{D}|-\kappa}\}$*, for* $j = 1, \ldots, \kappa$*. Then the collection of test sets* $\mathbb{K}$ *in Theorem A.2 can be initialized to sets of size* $|K| = m$*,* $m = 1, \ldots, |\mathcal{D}| - 1$*, comprised of adjacent alternatives (with respect to* $|\mathcal{D}|$*).*

Proof. For any non-empty set $K \subset \mathcal{D}$, $\Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K \neq \emptyset;\boldsymbol{\gamma}) = 1 - \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset K^C;\boldsymbol{\gamma})$, and therefore

$$\Pr(d \in K|\mathbf{x}) \leqslant \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \cap K \neq \emptyset;\boldsymbol{\gamma}) \\ \Leftrightarrow \Pr(d \in K^C|\mathbf{x}) \geqslant \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset K^C;\boldsymbol{\gamma}). \quad \text{(A.3)}$$

If $K = \{c_j, c_m\}$, then $K^C = \{c_1, \ldots, c_{j-1}, c_{j+1}, \ldots, c_{m-1}, c_{m+1}, \ldots, c_{|\mathcal{D}|}\}$, and

$$\Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset K^C;\boldsymbol{\gamma}) = \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset \{c_1, \ldots, c_{j-1}\};\boldsymbol{\gamma}) \\ + \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset \{c_{j+1}, \ldots, c_{m-1}\};\boldsymbol{\gamma}) + \Pr(D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\boldsymbol{\delta}) \subset \{c_{m+1}, \ldots, c_{|\mathcal{D}|}\};\boldsymbol{\gamma})$$

due to the structure of the realizations of $D^*_\kappa(\mathbf{x},\boldsymbol{\nu};\delta)$. Hence, due to the additivity of probabilities, the inequality in equation (A.3) for $K = \{c_j, c_m\}$ is satisfied whenever it holds for $K_1 = \{c_1, \ldots, c_{j-1}\}$, $K_2 = \{c_{j+1}, \ldots, c_{m-1}\}$, and $K_3 = \{c_{m+1}, \ldots, c_{|\mathcal{D}|}\}$, so that the inequality for $K = \{c_j, c_m\}$ is redundant. The same reasoning extends to any set K comprised of q alternatives, $q = 3, \ldots, |\mathcal{D}| - 1$, that are not all adjacent. □

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.2: *Let $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 3$. Suppose that all assumptions in Corollary 3.1 hold and that 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.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_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.2–A.3 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.

B Statistical Inference

The sample moments that we use to obtain the confidence intervals for (projections of) $\boldsymbol{\theta}$ in Section 5 are of the form:

$$\bar{m}_{n,K,j}(\boldsymbol{\theta}) = \widehat{\Pr}(d_i \in K | (\mu_i, \mathbf{p}_i) \in B_j) - \int_{B_j} P(D_\kappa^*(\mu, \mathbf{p}) \cap K \neq \emptyset; \boldsymbol{\gamma}) d\mu d\mathbf{p}, \tag{B.1}$$

where

$$\widehat{\Pr}(d_i \in K | (\mu_i, \mathbf{p}_i) \in B_j) = \frac{\sum_{i=1}^n \mathbf{1}(d_i \in K, (\mu_i, \mathbf{p}_i) \in B_j)}{\sum_{i=1}^n \mathbf{1}((\mu_i, \mathbf{p}_i) \in B_j)},$$

and the integral in equation (B.1) is computed using numerical approximation.

We obtain confidence regions based on the procedure proposed by Andrews and Soares (2010), as for example in Figure 5.1, and confidence intervals based on 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.

We base our confidence sets on the maximum moment violation statistic (function S_3 in Andrews and Soares (2010, p. 127)):

$$T_n(\boldsymbol{\theta}) = n \max_{j=1,\dots,64; K \in \mathbb{K}} \max \left\{ \frac{\bar{m}_{n,K,j}(\boldsymbol{\theta})}{\hat{\sigma}_{n,K,j}}, 0 \right\}^2$$

with $\hat{\sigma}_{n,K,j}$ the sample analog estimator of the asymptotic standard deviation of $\widehat{\Pr}(d_i \in K | (\mu_i, \mathbf{p}_i) \in B_j)$. Our application of the method proposed by Andrews and Soares (2010) computes bootstrap-based critical values to obtain a confidence set

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

with the property that it covers each $\boldsymbol{\theta} \in \Theta_I$ with asymptotic probability $1 - \alpha$ uniformly over a large class of probability distributions $\mathcal{P}$ described in Andrews and Soares (2010). Formally,

$$\liminf_{n \to \infty} \inf_{\mathsf{P} \in \mathcal{P}} \inf_{\boldsymbol{\theta} \in \Theta_I} \mathsf{P}(\boldsymbol{\theta} \in CS) \geqslant 1 - \alpha. \tag{B.2}$$

We use this method to compute a confidence set on $\boldsymbol{\gamma} = [\gamma_1, \gamma_2] \in \Gamma \subset \mathbb{R}^2$ (recalling that $\pi_3 = 1$ and $\pi_4 = \pi_5 = 0$ when projecting Θ_I on $\boldsymbol{\gamma}$).

In practice, we evaluate $T_n(\boldsymbol{\theta})$ and the bootstrap-based critical value $\hat{c}_{n,1-\alpha}(\boldsymbol{\theta})$ on a grid of values over $\Gamma = [0.01, 10] \times [0.01, 75.01]$ to obtain a precise description of CS. Our grid includes 1,501,000 points, with a step size of 0.01 on γ_1 and 0.05 on γ_2. The approximation of $\hat{c}_{n,1-\alpha}(\boldsymbol{\theta})$ is based on the bootstrap procedure detailed in Andrews and Soares (2010, Section 4.2) and uses 1,000 bootstrap replications.[37] The procedure takes as inputs a *GMS function* φ and a *GMS sequence* τ_n, 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 hard-threshold GMS function proposed by Andrews and Soares (2010):[38]

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

and we set $\tau_n = \sqrt{\log n}$ as recommended by Andrews and Soares (2010, Equation (4.4)).

We obtain confidence intervals on π_3, π_4, π_5, $E(\nu)$, $Var(\nu)$, and on the welfare cost of limited choice sets using the method proposed by Kaido et al. (2019). The first three parameters are linear projections of $\boldsymbol{\theta} = [\boldsymbol{\pi}, \boldsymbol{\gamma}]$. The other three are smooth functions of $\boldsymbol{\gamma}$ with gradients that satisfy the assumptions in Kaido et al. (2019, Theorem 3.1). 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,

$$\min_{\boldsymbol{\theta}\in\Theta} / \max_{\boldsymbol{\theta}\in\Theta} f(\boldsymbol{\theta}) \text{ s.t. } \sqrt{n}\bar{m}_{n,K,j}(\boldsymbol{\theta})/\hat{\sigma}_{n,K,j} \leqslant \hat{c}_n^f(\boldsymbol{\theta}),\ j = 1, ..., 64,\ K \in \mathbb{K},$$

where $\hat{c}_n^f(\boldsymbol{\theta})$ is computed using the bootstrap-based calibrated projection procedure detailed in Kaido et al. (2019, Section 2.2). The critical level $\hat{c}_n^f(\boldsymbol{\theta})$ is calibrated so that the function of $\boldsymbol{\theta}$, rather than $\boldsymbol{\theta}$ itself as in equation (B.2), is uniformly asymptotically covered with probability $1 - \alpha$ over a large class of probability distributions $\mathcal{P}$ described in Kaido et al. (2019). Formally,

$$\liminf_{n\to\infty} \inf_{\mathsf{P}\in\mathcal{P}} \inf_{\boldsymbol{\theta}\in\Theta_I} \mathsf{P}(f(\boldsymbol{\theta}) \in CI) \geqslant 1 - \alpha.$$

The procedure takes as inputs a GMS function φ and a GMS sequence τ_n, following Andrews and Soares (2010), for which we make the same choices as described above. The procedure

[37] Compared to the description in Andrews and Soares (2010, Section 4.2), note that our moment inequalities are of the $\leqslant$ form, whereas Andrews and Soares's are of the $\geqslant$ form.

[38] This is the function that they label $\varphi^{(1)}$ on p. 131. 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.

also requires a regularization parameter $\rho \geqslant 0$, which (similarly to φ and τ_n) enters the calibration of $\hat{c}^f_{n,1-\alpha}$ and introduces a conservative distortion that is required to obtain uniform coverage of projections. The smaller is the value of ρ, the larger is the conservative distortion, but the higher is the confidence that the critical value is uniformly valid in situations where the local geometry of Θ_I makes inference especially challenging. For a discussion, see Kaido et al. (2019, Section 2.2). We choose the value of ρ as follows. We begin with the recommendation in Kaido et al. (2019, 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.

References

Abaluck, J. and A. Adams (2018): "What Do Consumers Consider Before They Choose? Identication from Asymmetric Demand Responses," Working Paper, School of Management, Yale University.

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

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 (2019): "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.

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.

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 (2019): "A Random Attention Model," Working Paper, Department of Economics, University of Maryland.

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.

Ciliberto, F. and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

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.

Crawford, G. S., R. Griffith, and A. Iaria (2019): "Preference Estimation with Unobserved Choice Set Heterogeneity using Sufficient Sets," Discussion Paper 11675, Center for Economic Policy Research.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2018): "Inferring Cognitive Heterogeneity from Aggregate Choices," Working Paper Series No. 10-2018, Department of Economics, University of Sussex.

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.

Gabaix, X. (2018): "Behavioral Inattention," Working Paper, Department of Economics, Harvard University.

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.

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.

Harless, D. and C. Camerer (1994): "The Predictive Utility of Generalized Expected Utility Models," *Econometrica*, 62, 1251–1289.

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 (2016): "Inattention and Switching Costs as Sources of Inertia in Medicare Part D," Working Paper 22765, National Bureau of Economic Research.

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.

Horowitz, J. L. and J. J. Louviere (1995): "What is the Role of Consideration Sets in Choice Modeling?" *Interantional Journal of Research in Marketing*, 12, 39–54.

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.

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

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

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.

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. (2018): "Estimating Multinomial Choice Models with Unobserved Choice Sets," Working Paper, Shanghai University of Finance and Economics.

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.

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.

Matzkin, R. L. (2007): "Nonparametric Identification," in *Handbook of Econometrics,* Vol. 6B, ed. by J. J. Heckman and E. E. Leamer, Amsterdam: North-Holland, 5307–5368.

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. (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.

Molinari, F. (2019): "Econometrics with Partial Identification," CeMMAP Working Paper CWP25/19.

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.

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

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.

Train, K. E. (2009): *Discrete Choice Methods with Simulation* (Second Ed.), Cambridge: Cambridge University Press.

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.

Wilcox, N. T. (2008): "Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison," in *Risk Aversion in Experiments*, ed. by J. C. Cox and G. W. Harrison, Bingley: Emerald, 197–292.

Supplement to "Heterogeneous Choice Sets and Preferences"

Levon Barseghyan
Cornell University
lb247@cornell.edu

Maura Coughlin
Cornell University
mc2463@cornell.edu

Francesca Molinari
Cornell University
fm72@cornell.edu

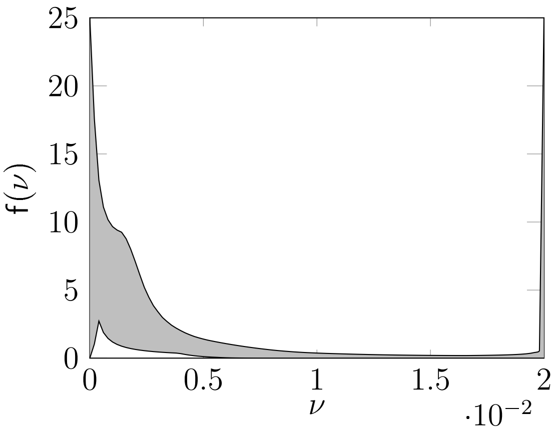
Joshua C. Teitelbaum
Georgetown University
jct48@georgetown.edu

June 30, 2019

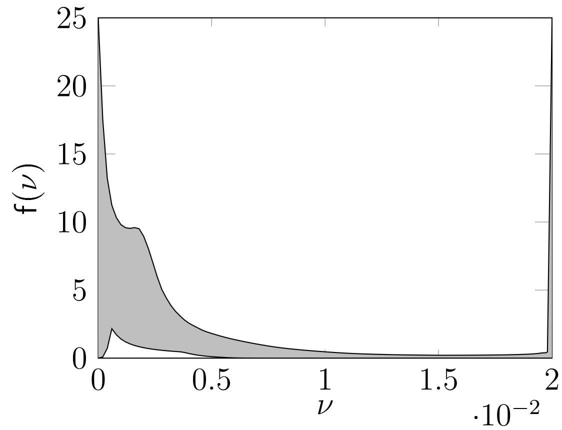
Table S.1: Auto Collision Claim Rate Regression

Variable	Coefficient	Standard error
Constant	−6.6768	0.0761
Driver 2 indicator	0.2389	0.0486
Driver 3+ indicator	0.5585	0.0630
Vehicle 2 indicator	0.4362	0.0478
Vehicle 3+ indicator	0.5972	0.0592
Young driver	0.1028	0.0253
Driver 1 age	−0.0182	0.0014
Driver 1 age Squared	0.0002	0.0000
Driver 1 female	0.0441	0.0085
Driver 1 married	0.0694	0.0099
Driver 1 divorced	0.0663	0.0130
Driver 1 separated	0.0970	0.0229
Driver 1 single	.	.
Driver 1 widowed	0.0498	0.0149
Vehicle 1 age	−0.0433	0.0015
Vehicle 1 age squared	0.0008	0.0001
Vehicle 1 business	.	.
Vehicle 1 farm	−0.2366	0.0873
Vehicle 1 pleasure	−0.1171	0.0284
Vehicle 1 work	−0.1039	0.0283
Vehicle 1 passive restraint	−0.0826	0.0263
Vehicle 1 anti-theft	0.0180	0.0074
Vehicle 1 anti-break	−0.0080	0.0078
Driver 2 age	0.0037	0.0021
Driver 2 age squared	0.0000	0.0000
Driver 2 female	0.0678	0.0134
Driver 2 married	−0.2062	0.0201
Driver 2 divorced	−0.1382	0.0851
Driver 2 separated	−0.2019	0.1777
Driver 2 single	.	.
Driver 2 widowed	−0.2601	0.1291
Vehicle 2 age	−0.0308	0.0016
Vehicle 2 age squared	0.0005	0.0001
Vehicle 2 business	.	.
Vehicle 2 farm	−0.2683	0.1131
Vehicle 2 pleasure	−0.1591	0.0361
Vehicle 2 work	−0.1619	0.0362
Vehicle 2 passive restraint	0.0237	0.0248
Vehicle 2 anti-theft	0.0342	0.0098
Vehicle 2 anti-break	0.0107	0.0102
Insurance score	−0.0018	0.0000
Previous accident	0.0827	0.0147
Previous convictions	0.1336	0.0862
Previous reinstated	0.0354	0.0515
Previous revocation	−0.1037	0.1451
Previous suspension	0.0434	0.0521
Previous violation	0.0953	0.0086
Year dummies	Yes	
Territory codes	Yes	
Variance (ϕ)	0.1733	0.0057
Loglikelihood	-426,901	

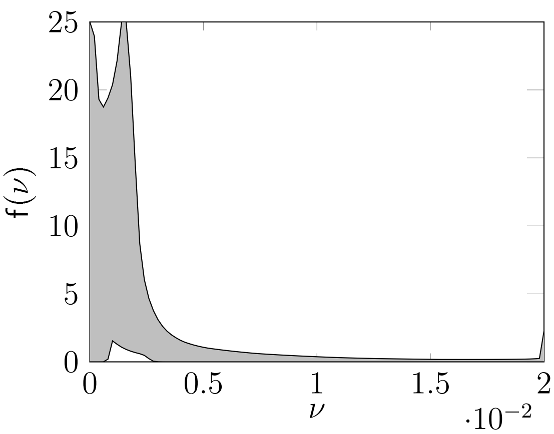
Notes: Poisson panel regression with random effects (1,349,853 observations). Insurance score is a credit based risk score. Territory codes indicate rating territories, which are based on actuarial risk factors such as traffic and weather patterns, population demographics, wildlife density, and the cost of goods and services.



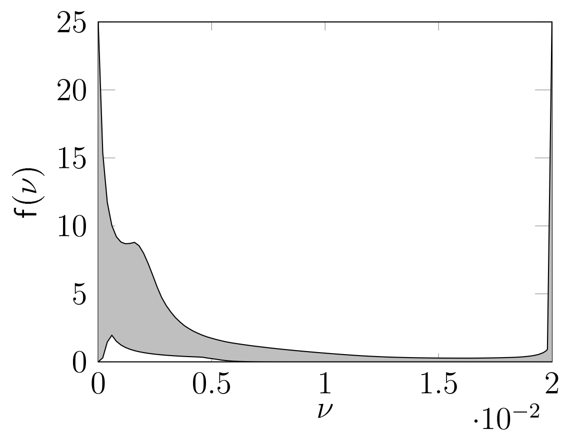
(a) Male



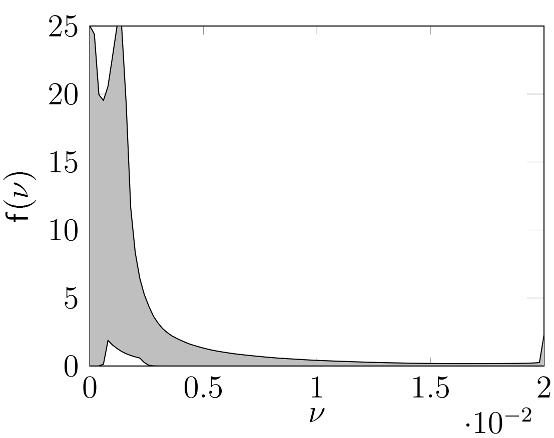
(b) Female



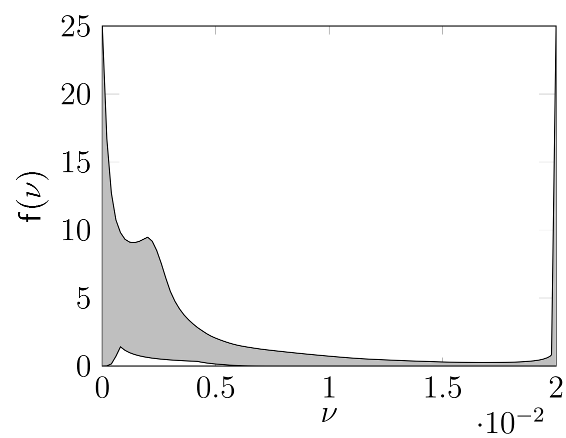
(c) Young



(d) Old



(e) Low insurance score



(f) High insurance score

Figure S.1: Outer region of admissible probability density functions of ν.

Note: The figure depicts the outer region of admissible probability density functions of ν_i based on the AS confidence set for $\boldsymbol{\theta}$ for selected subsamples based on gender, age, and insurance score. Insurance score is a credit based risk score. 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.