ESSAYS ON CONSUMER CHOICE WITH UNOBSERVED CHOICE SETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Maura Coughlin

May 2020

ESSAYS ON CONSUMER CHOICE WITH UNOBSERVED CHOICE SETS

Maura Coughlin, Ph.D.

Cornell University 2020

This dissertation consists of three essays that evaluate how consumers make decisions in settings where the researcher may not know the set of alternatives from which observed choices were selected. Many empirical analyses in economics presume the researcher knows the full set of alternatives an individual compared when selecting their most preferred. In practice, this assumption may fail to hold for a variety of reasons. In the first chapter, I introduce the economic setting of unobserved choice sets and consideration sets defining to this work.

In the second chapter, my coauthors and I 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 then apply our theoretical findings to learn about households' risk preferences and choice sets from data on their deductible choices in auto collision insurance.

The third chapter evaluates the prescription drug insurance choices of Medicare beneficiaries. I propose an empirical model of demand for prescription drug plans where non-monetary plan attributes stochastically determine the composition of the set of plans that an individual considers, and monetary plan attributes determine the individual's expected utility over contracts in her consideration set. This model reconciles the classic view of insurance contracts as

lotteries with purely monetary outcomes with the empirical finding that choice among insurance plans is driven by their non-monetary attributes and financial attributes beyond their impacts on costs. I estimate the model using data from Medicare Part D allowing for unobserved heterogeneity in risk aversion and in consideration sets. I find that the latter plays a crucial role in plan choices, and in contrast to previous literature that assumes full consideration of all plans, I uncover an important role for risk aversion in determining individual choices.

BIOGRAPHICAL SKETCH

Maura Coughlin grew up in Mundelein, Illinois. She attended Johns Hopkins University, graduating with majors in economics and international studies. Prior to pursuing a Ph.D. in Economics at Cornell University, she worked as a research analyst at The Brattle Group in Washington, D.C. After graduation, she is joining the Economics Department at Rice University as an Assistant Professor.

To my loving and supportive

Coughlin, O'Hara, and Mogen families.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of my committee, and any acknowledgment here will surely fail to adequately thank them. My committee chairs, Levon Barseghyan and Francesca Molinari, were endlessly generous with their time throughout my graduate career. Without their guidance, this work would have been unimaginable. Their encouragement has been pivotal in my professional development and personal growth, and I cannot thank them enough for all they have taught me over these years. Colleen Carey kindly shared her data and provided crucial feedback for the last chapter of my dissertation. Her generosity with her time, knowledge, and data saved me an immeasurable amount of time and hassle.

I would also like to thank my co-author Joshua Teitelbaum, who along with Levon Barseghyan and Francesca Molinari worked with me to bring the second chapter to life. Working alongside them provided invaluable education in the creativity, rigor, and determination behind research that I will take forward in my career. My research and graduate experience have also benefited from many conversations with the wonderful faculty at Cornell. I would also like to thank my fellow graduate students, whose assistance helped me through first year qualifying exams and the job market and whose friendships have made Ithaca home. I would especially like to thank Lin Xu for her help on all things data, Barton Willage for his mentorship, and Matthew Thirkettle for being such an excellent officemate. I would, of course, like to thank my family for their support through graduate school and the years leading here. This work was supported in part by the Institute for Social Sciences at Cornell University.

Above all, and for everything, I thank Philip Mogen.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CONSUMER CHOICE WITH UNOBSERVED CHOICE AND CONSIDERATION SETS: AN OVERVIEW

In traditional discrete choice models popularized by McFadden (1974), estimation of unobserved preferences relies on the principle of revealed preference. Given a finite set of alternatives from which to choose, individuals are presumed to select their most preferred option. Such an assumption paired with a model of utility allows researchers to learn about consumer valuations of alternatives as the utility of the chosen alternative is revealed to exceed that of the available but declined alternatives. This modeling assumption fundamentally rests on the researcher's ability to characterize the set of theoretically available alternatives, or more precisely, the set of alternatives over which an individual compared utility. In practice, however, this assumption is often unverifiable. In other cases, there is reason to suspect that consumers are not selecting from a full set of options observable to or describable by the researcher.

My dissertation explores the issue of estimating consumer demand in the presence of unobserved choice sets in more detail. In one chapter, coauthors and I propose a new theoretical model of discrete choice robust to unobserved choice sets. This model does not require the researcher to have additional information on how choice sets arise from the set of feasible alternatives, and while partially rather than point identified, can provide informative bounds on parameters of interest. In my other chapter, I employ a specific model of limited consideration to the choice of standalone prescription drug plans among Medicare beneficiaries. The specific empirical setting of that chapter is one in which the assumption that consumers evaluate all alternatives is rather suspect.

To obtain point identification of a model with unobserved choice sets, additional assumptions are needed to complete the model. Frequently - especially in the cases where there is not an explicit discussion of the observability of choice sets - the assumption is made that each individual's choice set corresponds to the entire feasible set of alternatives. Other models of heterogeneous choice sets, including the model employed in my final chapter, place structure on how choice sets form from the feasible set and leverage additional information or additional restrictions to point identify preferences. Erroneously assuming such a point identified model can result in incorrect estimates, an issue explored in the empirical portion of my coauthored chapter.

In practice the standard modeling assumption that the choice sets individuals evaluate corresponds to the entire feasible set can fail for a variety of plausible reasons. A simple example is product unavailability. If a consumer purchasing a good finds their preferred product out of stock, they may simply select their next most preferred product, and the researcher is unaware that the most preferred option was not available at the time of purchase. In other settings unobserved constraints, such as liquidity or budget constraints, may limit what subset of theoretically feasible alternatives are feasible in practice for a given consumer. Although an individual's preferences may suggest a luxury automobile delivers higher utility than other vehicles, a consumer may deem it unaffordable and therefore not evaluate it when purchasing an automobile. Such a consumer behaves in a fully rational manner, but the unobservability of the constraints binding their choices could lead a researcher to incorrect inference. Others consumers may exhibit limited consideration, where they (for a variety of reasons) do not consider all feasible alternatives, possibly due to cognitive constraints or due to effects of advertising or agent effects.

In my coauthored chapter entitled "Heterogeneous Choice Sets and Preferences," we propose a robust discrete choice model when choice sets are unobserved. The model relies on a small number of (relative to many existing approaches) weak assumptions about what the researcher observes about the choices individuals make. Specifically, aside from assuming a nontrivial minimum size of choice sets, our model does not impose restrictions or unverifiable assumptions about individuals' choice sets. While the researcher must have sufficient information to list the set of feasible superset of alternatives, the model allows for unobserved choice sets to take any form and have arbitrary correlation with individual preferences. As such, the model is consistent with many underlying models of choice set formation. Thus, it does not require the researcher to have additional data to specify that formation process and avoids concerns of misspecification arising from imposing a specific mechanism.

Due to the agnostic nature of the model with regards to choice sets, identical individuals may make different rational choices if different choice sets manifest. For example, an individual whose choice set includes the option that delivers the highest utility from the feasible set will select that first best option. If the choice set does not contain the first best option but does include the second best option, then the optimal choice would be to select the second best alternative. Thus, for any individual with a given value of preferences, there are multiple "optimal" choices. The set of model implied optimal choices is, therefore, a random set. We establish that this model of discrete choice is partially identified and use random set theory to characterize the sharp identification region which collects the set of model primitives consistent with the observed choices and modeling assumptions.

We apply our model to learn about households' risk preferences and choice sets using data on collision deductible insurance. In this specific setting the observed choice patterns are inconsistent with many standard models of discrete choice and the assumption that individuals made their choices from the full set of feasible deductibles. The bounds on risk preferences and consideration are informative and suggest that observed choices can be rationalized with lower and more homogeneous risk aversion than suggested in standard models assuming choices are made evaluating the entire feasible set. This method is applicable broadly when choice sets are unobserved and either researchers do not want to enforce or the data reject the assumptions needed to garner point identification.

While the model described above is robust to arbitrary choice set formation processes, the estimation focuses on recovering preferences while saying as little as necessary about choice sets, aside from possible size. In other settings, the process through which the feasible choice sets translates into a consideration set that is of particular interest as well. In my last chapter, "Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D," I estimate both risk preferences and the consideration behavior of Medicare beneficiaries choosing standalone prescription drug insurance plans. As in many health insurance markets, individual choice over Medicare Part D prescription drug plans are difficult to rationalize with standard models of discrete choice. Individuals frequently select plans that with substantially higher mean and variance of out-of-pocket costs for their drug needs than other available insurance plans. Such a patterns is not explained by heterogeneous risk aversion alone. Moreover, attributes of the insurance plans economists generally do not consider relevant for evaluating and comparing insurance products (such as the identity of the insuring firm) have strong relationships with choice proba-

bilities. Taken together, these patterns documented extensively in the literature are suggestive that observed choice behavior in this market does not conform with standard models of insurance demand.[1] Such a model, however, makes the implicit assumption that individuals are making plan enrollment decisions from the entire set of available plans.[2] In the early years of the Medicare Part D program, the typical beneficiary faced the choice of around 50 different complex plans, and many researchers have suspected that individuals relied on heuristic shortcuts to make a choice from such a large and complex choice environment. In my dissertation chapter, I propose an alternative model of insurance demand that formalizes this intuition.

The approach to demand estimation differs in different contexts. Some consumer products can be viewed as a collection of attributes, and individuals are assumed to have preferences over those attributes. In other settings, however, there is a foundation of economic theory underlying demand. This is true, in particular, in insurance markets. Theory of expected utility suggests that insurance appeals to risk averse individuals facing uncertainty in their future as a means of transferring resources (money) from potential good states of the world (where the uncertainty resolves favorably) to bad states of the world (where uncertainty resolves in a very costly manner). In previous studies, where models approximating expected utility under the assumption of full consideration appear inconsistent with empirical choice patterns, reduced form models of demand over product attributes can rationalize choice data but in a manner that is both challenging to interpret and abstracts from the structural role of risk aversion in insurance demand.

[1] See, for example, Abaluck and Gruber (2011), Ketcham et al. (2016), Abaluck and Gruber (2016a), Ketcham et al. (2012), and Heiss et al. (2013).

[2] Exceptions to this assumption in the Medicare Part D literature focus specifically on the topic of inertia, such as in Ho et al. (2017) and Abaluck and Gruber (2016b).

In settings, such as prescription drug insurance, where a theoretical foundation of demand is established, empirical patterns suggest additional product attributes are at play, and the feasible choice set is considered challengingly large, limited consideration offers a natural way to account for the role of such product attributes outside the model by allowing those attributes to stochastically determine the set of plans an individual considers while maintaining the assumption of the structural utility model. In this chapter I propose a model of insurance demand that marries classic expected utility with an alternative specific consideration model. In contrast to the previous literature, I find risk preferences that are very similar to what has been estimated in similar insurance contexts. Heterogeneity in consideration plays a crucial role in explaining plan choices among Medicare beneficiaries, and consideration is largely driven by highly visible plan attributes.

The choice patterns this chapter addresses have sparked interest from other economists offering alternative models to insurance choice. To model rational behavior that results in beneficiaries choosing financially dominated plans, Brown and Jeon (2019) use a model of rational inattention, whereby individuals can easily learn an insurance plan's premium but must exert effort to learn about the plan's out-of-pocket drug cots. This study shares some motivation with my dissertation chapter, but is empirically motivated by a reduced form data pattern wherein individuals for whom the stakes are higher seek more information and choose better plans. This pattern does not materialize in my data sample. Others have proposed models of multiple types of agents with informed agents whose choices reveal preferences and uninformed agents who are not choosing optimally, such as in Ketcham et al. (2019) and Keane et al. (2019). My dissertation along with these and similar studies in other markets,

demonstrate a movement towards relaxing standard modeling assumptions to more realistically reflect the way individuals make economic decisions.

Individual choices recorded in economic data reflect the interaction of consumer preferences and the sets of alternatives from which individuals made selections. In many applications, it is difficult for a researcher to know the composition of that set of alternatives. My dissertation reflects the shift of the field towards relaxing the at times unrealistic assumption that individuals evaluate and compare all alternatives feasibly available to them. Inference about consumer preferences can be sensitive to assumptions regarding individual choice sets. The empirical results in my dissertation highlight that reality, while the theoretical contributions offer alternative approaches to empiricists.

CHAPTER 2

HETEROGENEOUS CHOICE SETS AND PREFERENCES (WITH LEVON BARSEGHYAN, FRANCESCA MOLINARI, AND JOSHUA C. TEITELBAUM)

2.1 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 mis-

specified 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 2.2. 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 2.3 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 2.4 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 2.5. 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 para-

graph 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. (2016) 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 2.6 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.

2.2 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 2.2.1 (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 2.2.1(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 2.2.1(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 2.3; see Remark 2.3.1.

REMARK 2.2.1: 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 2.2.2 (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 2.2.2(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 2.2.2(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 2.2.2(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 2.2.2(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 2.2.2: 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.2.

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.

2.3 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 2.3.1). 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 2.3.2). 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 2.3.3).

2.3.1 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 2.2 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

$$D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) = \bigcup_{G \subseteq \mathcal{D}: |G| \geqslant \kappa} \Big\{ d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \Big\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \Big\{ d_i^*(G; \mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}) \Big\}, \quad (2.2)$$

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 2.1 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

[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.2, Definition A.2.1. That $D_\kappa^*(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ is a random closed set is formally established in Appendix A.2, Lemma A.2.1.

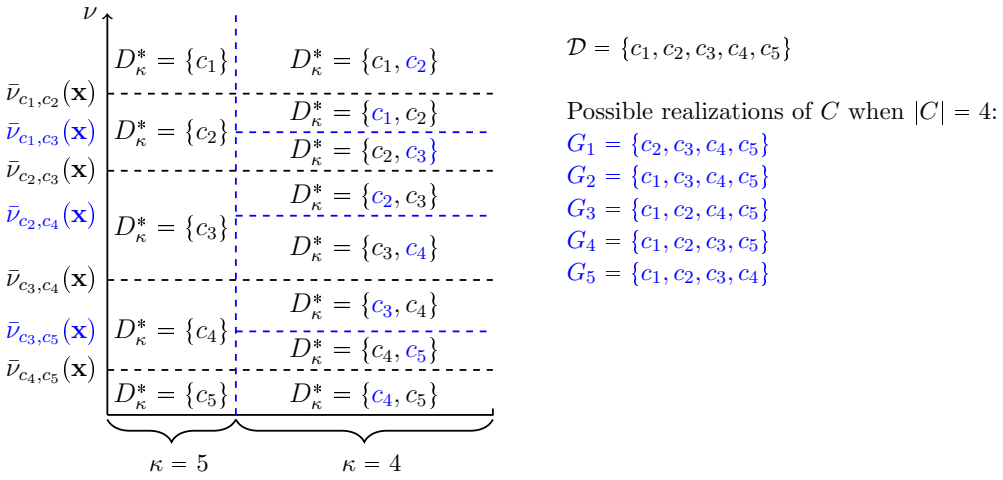


Figure 2.1: Stylized depiction of D^*_κ when $|\mathcal{D}| = 5$ and $\kappa \in \{4, 5\}$.

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 $\kappa \in \{4, 5\}$.

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 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 2.2 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 differ-

ent 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 $c \in \mathcal{D}$,

$$\Pr(d_i^* = c|\mathbf{x}_i; \boldsymbol{\theta}) = \int_{\boldsymbol{\tau} \in \mathcal{V}} \sum_{G \subseteq \mathcal{D}} \mathbf{1}(d_i^*(G; \mathbf{x}_i, \boldsymbol{\tau}; \boldsymbol{\delta}) = c) \mathsf{F}(G; \mathbf{x}_i, \boldsymbol{\tau}) dP(\boldsymbol{\tau}; \boldsymbol{\gamma}). \tag{2.3}$$

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

$$\Pr(d_i^* = c|\mathbf{x}_i; \boldsymbol{\theta}) = \Pr(d_i = c|\mathbf{x}_i), \ \forall c \in \mathcal{D}, \ \mathbf{x}_i - a.s., \tag{2.4}$$

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 (2.4), 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 2.1. 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 (2.4) 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 dimen-

[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 2.1 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.

sional 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

$$d_i \in D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta}),\ \mathbf{x}_i - a.s. \tag{2.5}$$

for the data generating value of $\boldsymbol{\theta}$. To harness the empirical content of equation (2.5) 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.2, Theorem A.2.1. This result allows us to translate equation (2.5) 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]

THEOREM 2.3.1: *Let Assumptions 2.2.1 and 2.2.2 hold and let $\Theta = \Delta \times \Gamma$. Then*

$$\Theta_I = \left\{ \boldsymbol{\theta} \in \Theta : \Pr(d \in K|\mathbf{x}) \leqslant P(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \forall K \subseteq \mathcal{D}, \mathbf{x} - a.s. \right\}. \tag{2.6}$$

It is immediate that if equation (2.5) holds then the inequalities in equation (2.6) 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 (2.6) are satisfied for each $K \subseteq \mathcal{D}$ then equation (2.5) holds. Our proof of Theorem 2.3.1, provided in

[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).

Appendix A.2, establishes that the characterization in equation (2.6) is sharp—all and only those values of $\boldsymbol{\theta}$ for which the inequalities in equation (2.6) 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 (2.6), $\Pr(d_i \in K|\mathbf{x}_i)$, 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.2, Theorem A.2.2 we provide an algorithm that substantially reduces the number of inequalities that need to be checked to obtain Θ_I.

REMARK 2.3.1: Theorem 2.3.1, as well as Corollary 2.3.1 and Theorem 2.3.2 below, can be generalized for a *structure* (W, P) (or (W, P, π) in the case of Corollary 2.3.1) 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.

2.3.2 Preferences and Choice Sets

Theorem 2.3.1 establishes that, under mild restrictions on the utility function (Assumption 2.2.1) and knowing only the minimum size of agents' choice sets (Assumption 2.2.2), 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 2.2.1 and 2.2.2, we now assume that:

ASSUMPTION 2.3.1 (Choice Set Size): *Agent i draws the size ℓ_i of her choice set such that*

$$\Pr(\ell_i = q|\mathbf{x}_i, \boldsymbol{\nu}_i) = \Pr(\ell_i = q|\mathbf{x}_i) = \pi(q; \mathbf{x}_i; \boldsymbol{\eta}),\ q = \kappa, \ldots, |\mathcal{D}|, \tag{2.7}$$

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 2.3.1 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 2.3.1, Theorem 2.3.1 specializes to the following corollary.[6]

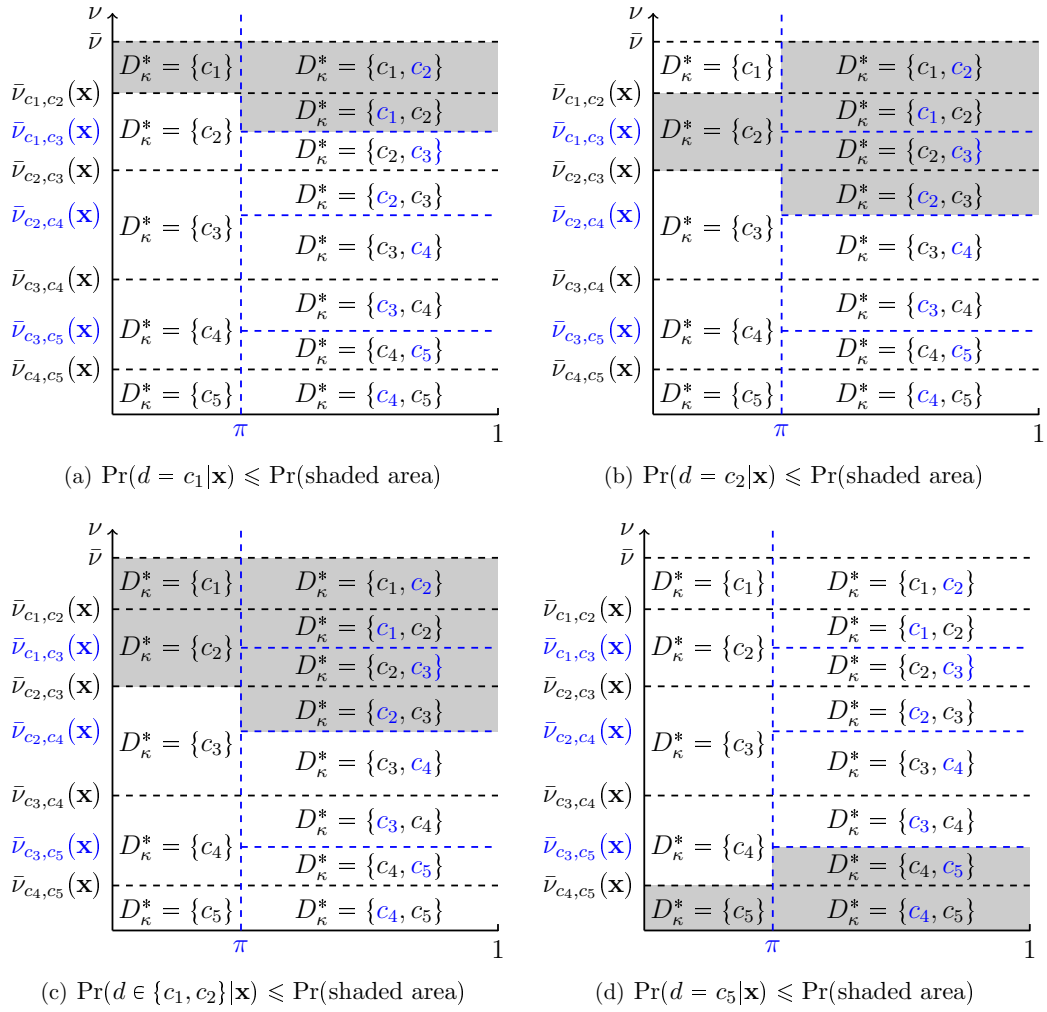
[6]The proof of Corollary 2.3.1 follows immediately from the proof of Theorem 2.3.1 and therefore is omitted.

COROLLARY 2.3.1: *Let Assumptions 2.2.1, 2.2.2, and 2.3.1 hold and let* $\boldsymbol{\theta} = [\boldsymbol{\eta}; \boldsymbol{\delta}; \boldsymbol{\gamma}]$ *and* $\Theta = H \times \Delta \times \Gamma$. *Then*

$$\Theta_I = \Bigg\{ \boldsymbol{\theta} \in \Theta : \Pr(d \in K | \mathbf{x}) \leqslant \sum_{q=\kappa}^{|\mathcal{D}|} \pi_q(\mathbf{x}; \boldsymbol{\eta}) P(D_q^*(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \quad \forall K \subseteq \mathcal{D}, \mathbf{x} - a.s. \Bigg\}. \tag{2.8}$$

The sharp identification region Θ_I in Corollary 2.3.1 has two noteworthy features. First, the projection of Θ_I on $[\boldsymbol{\delta}; \boldsymbol{\gamma}]$ is equal to the sharp identification region in Theorem 2.3.1. In other words, the information in Θ_I about the distribution of preferences is the same with or without Assumption 2.3.1. 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 2.2 contains stylized depictions of selected inequalities in equation (2.8) 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^* 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



(a) $\Pr(d = c_1|\mathbf{x}) \leqslant$ Pr(shaded area)

(b) $\Pr(d = c_2|\mathbf{x}) \leqslant$ Pr(shaded area)

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

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

Figure 2.2: Stylized depictions of selected inequalities in Θ_I when $|\mathcal{D}| = 5$ and $\kappa = 4$.

Note: The figure depicts the inequalities in equation (2.8) 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) $K = \{c_5\}$.

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 (2.8) 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 2.3.1) and is formalized in Appendix A.2, Theorem A.2.2.

Though not depicted in Figure 2.2, 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 2.4 and 2.5 we apply Theorem 2.3.1 and Corollary 2.3.1 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 2.4. We present our empirical findings in Section 2.5. 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. (2016) to obtain asymptotically uniformly valid confidence intervals for smooth functions and components of $\boldsymbol{\theta}$.[7]

[7]For details on our empirical methods, see Section 2.5.1 and Appendix A.3.

2.3.3 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.

THEOREM 2.3.2: *Let Assumptions 2.2.1 and 2.2.2 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*

$$
\begin{aligned}
&\max_{\boldsymbol{\theta}\in\Theta} \ \mathrm{E}(W(\mathbf{x}_{d^{|\mathcal{D}|}(\boldsymbol{\nu})}, \boldsymbol{\nu}; \boldsymbol{\delta}); \boldsymbol{\gamma}) - \mathrm{E}(W(\mathbf{x}_{d^{\kappa}(\boldsymbol{\nu})}, \boldsymbol{\nu}; \boldsymbol{\delta}); \boldsymbol{\gamma}) \qquad (2.9) \\
&\text{s.t.} \ \Pr(d \in K|\mathbf{x}) \leqslant P(D^*_{\kappa}(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}), \ \forall K \subseteq \mathcal{D}, \ \mathbf{x} - a.s.
\end{aligned}
$$

The proof of Theorem 2.3.2 follows immediately from Theorem 2.3.1. Indeed, Theorem 2.3.1 yields that Θ_I is characterized by the constraints in problem (2.9). 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 (2.9). Maximizing over Θ_I yields the result.

Under our assumptions, the objective function in problem (2.9) is smooth in the parameter vector $\boldsymbol{\theta}$. Therefore, we can apply the calibrated projection method proposed by Kaido et al. (2016) to obtain an asymptotically uniformly valid confidence interval for the solution to problem (2.9). For details, see Appendix A.3.

2.4 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 2.4.1 we specify a random expected utility model that allows for unobserved heterogeneity in risk aversion and choice sets and maintains Assumptions 2.2.1, 2.2.2, and 2.3.1. In Section 2.4.2 we describe our data. In Section 2.4.3 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 2.5.

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.

2.4.1 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 2.4.1 (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 2.4.1(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 2.4.1(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).

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

Under Assumption 2.4.1, household i's choice of deductible involves a choice among binary lotteries, indexed by $c \in \mathcal{D}$, of the following form:

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

The household chooses among these lotteries based on the criterion in equation (2.1).

We assume that household i's preferences conform to expected utility theory:

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

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 2.4.2 (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 2.2 and 2.3, 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 2.2.1 and 2.4.2, we posit that $\nu_i \sim P(\boldsymbol{\gamma}(\mathbf{t}_i))$, where P is specified

[9] In terms of the general notation used in Sections 2.2 and 2.3, $\mathbf{s}_i = (\mu_i, \mathbf{t}_i)$, $\mathbf{z}_{ic} = p_{ic}$, and $\boldsymbol{\nu}_i = \nu_i$.

below in Assumption 2.4.3(I), and that for all $c \in \mathcal{D}$,

$$W(\mathbf{x}_{ic}, \nu_i) = \frac{(1-\mu_i)(1-\exp(\nu_i p_{ic})) + \mu_i(1-\exp(\nu_i(p_{ic}+c)))}{\nu_i} \tag{2.10}$$

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

Observe that, by equation (2.10), 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 2.2.2(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 2.2.2(II) and 2.3.1, 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 (2.7). We could specify $\pi_q(\mathbf{t}_i) = f(\mathbf{t}_i; \boldsymbol{\eta})$ where the functional form of f is known up 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 2.4.3 (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 2.4.3(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 2.4.3(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 2.4.2, 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

$$D_\kappa^*(\mathbf{x}_i, \nu_i) = \bigcup_{G \subseteq \mathcal{D}: |G| \geqslant \kappa} \left\{ d_i^*(G; \mathbf{x}_i, \nu_i) \right\} = \bigcup_{G \subseteq \mathcal{D}: |G| = \kappa} \left\{ d_i^*(G; \mathbf{x}_i, \nu_i) \right\}. \tag{2.11}$$

The sharp identification region Θ_I of the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\pi}; \boldsymbol{\gamma}]$ is given by equation (2.8) where P is specified in Assumption 2.4.3(I) and D^*_κ is given by equation (2.11).

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 2.4.1

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

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 2.4.1.

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 2.2.2(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 2.4.3. As κ decreases between 4 and 2 the model gains 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

[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.

Section 2.5. 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 2.5. 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 2.4.1(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 vari-

ance 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 2.4.1(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.

Tables 2.1 and 2.2 present descriptive statistics for the analysis sample. Table 2.1 summarizes the households' deductible choices, pricing menus, claim probabilities, and demographic characteristics. Table 2.2 reports the sample distribution of deductible choices for the full sample and for selected subsamples based on gender, age, and insurance score. In Table 2.2 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 2.2 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

[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 2.1: 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.

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 2.2: 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: 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.

2.4.3 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

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

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.

[14] Evaluating $W(\mathbf{x}_{ic}, \nu_i)$ in equation (2.10) 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).

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 2.2.

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 2.4.1, including the assumption that households have expected utility preferences with Beta distributed, constant absolute risk aversion.

DEFINITION 2.4.1 (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 2.4.2 (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 2.4.1: *Take the model in Section 2.2. 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 2.4.1(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 2.4.1(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 2.4.1 (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* $W(\mathbf{x}_{ic'},\boldsymbol{\nu}_i;\boldsymbol{\delta})\geqslant W(\mathbf{x}_{ic},\boldsymbol{\nu}_i;\boldsymbol{\delta}),\ \boldsymbol{\nu}_i-a.s.$

[15] We could condition the choice set distribution on $\mathbf{t}_i$ and the result in Claim 2.4.1(II) would go through mutatis mutandis.

To see that MixL satisfies Property 2.4.1, note that under MixL,

$$\Pr(d_i = c' | \mathbf{x}_i, \boldsymbol{\nu}_i) = \frac{\exp(W(\mathbf{x}_{ic'}, \boldsymbol{\nu}_i; \boldsymbol{\delta}))}{\sum_{c \in \mathcal{D}} \exp(W(\mathbf{x}_{ic}, \boldsymbol{\nu}_i; \boldsymbol{\delta}))}. \tag{2.12}$$

Property 2.4.1 follows from equation (2.12) for any distribution of $\boldsymbol{\nu}_i$ with support $\mathcal{V}$, and Claim 2.4.1(I) follows from Property 2.4.1 by integrating with respect to the distribution of $\boldsymbol{\nu}_i$. Indeed, any discrete choice model that satisfies Property 2.4.1—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 2.4.1(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 2.2. The first, which we call the 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 2.4.3 (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 2.4.4 (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]

[16]We could condition the choice set distribution on $\mathbf{t}_i$ and the result in Claim 2.4.2(II) would

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 2.4.2: *Take the model in Section 2.2. 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 2.4.2(I) follows from the fact that our model coupled with UR satisfies Property 2.4.1. 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 2.4.2(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 $c' \in \mathcal{D}\backslash\{a, b\},\ c' \neq c$. More generally, $\Pr(d = a|\mathbf{x}) + \Pr(d = b|\mathbf{x}) < \Pr(d = c|\mathbf{x})$ is possible as long as $\varphi(a)$ and $\varphi(b)$ are sufficiently low, $\varphi(c)$ is sufficiently high, and c is the best alternative in $\mathcal{D}\backslash\{a, b\}$ for some positive measure of values of $\boldsymbol{\nu}_i \in \mathcal{V}$.

REMARK 2.4.1: 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

go through mutatis mutandis.

and represents expected utility 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.

2.5 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

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

more credible estimates of risk preferences.

We begin with a brief explanation of our empirical methods (Section 2.5.1). We then apply Theorem 2.3.1 to learn about the distribution of risk preferences (Section 2.5.2). After that, we apply Corollary 2.3.1 to learn about the distribution of choice set size (Section 2.5.3) and Theorem 2.3.2 to assess the welfare cost of limited choice sets (Section 2.5.4).

2.5.1 Summary of Empirical Methods

The inequalities in equations (2.6) and (2.8) and in the constraints in problem (2.9) need to hold $(\mu_i, \mathbf{t}_i, \mathbf{p}_i) - a.s.$ As indicated in Section 2.4.1, 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 bins $B_j,\ j = 1, \ldots, J$. 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.2, Theorem A.2.2 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

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

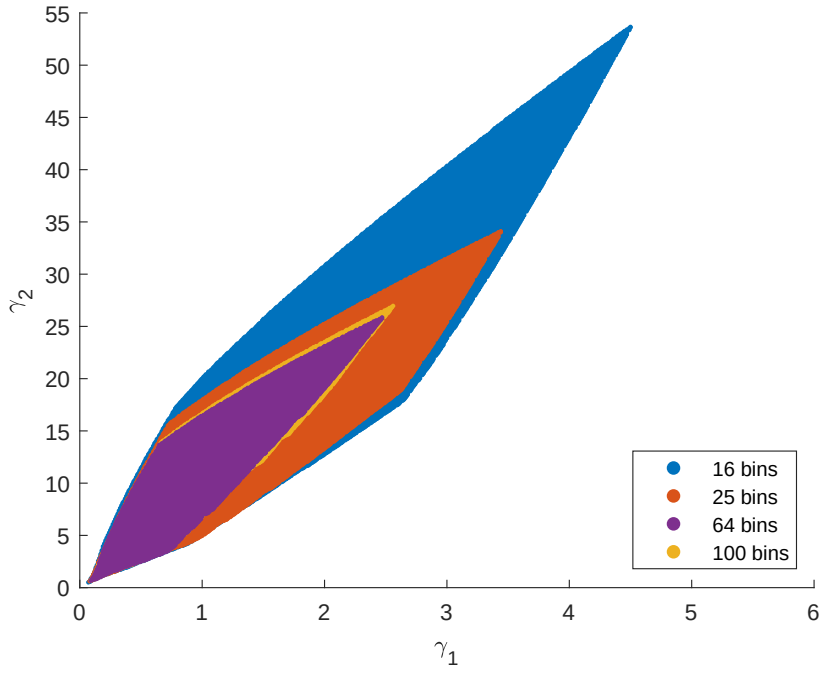


Figure 2.3: Confidence set for $\boldsymbol{\theta}$.

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.

and Soares (2010) [hereafter, AS] to obtain confidence sets that asymptotically uniformly cover $\boldsymbol{\theta} \in \Theta_I$ with probability 95 percent. Under Theorems 2.3.1 and 2.3.2, $\boldsymbol{\theta} = (\gamma_1, \gamma_2)$. Under Corollary 2.3.1, $\boldsymbol{\theta} = (\gamma_1, \gamma_2, \pi_3, \pi_4, \pi_5)$. We apply the bootstrap-based calibrated projection method proposed by Kaido et al. (2016) [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 A.3.

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 2.3. As we refine the discretized suppport, the resulting inequalities are different aggregations of the inequalities in equations (2.6), (2.8), and (2.9) 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 (2.6) and (2.8) and the constraints in problem (2.9) are satisfied. This implies that we fail to reject the hypothesis that our empirical model is 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).

Table 2.3: Risk Preferences and Welfare

	(1) $\mathrm{E}(\nu)$		(2) $\mathrm{Var}(\nu)$		(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 (UB) denotes lower (upper) bound. The lower bound on the welfare cost is zero by definition.

2.5.2 Risk Preferences

Columns (1) and (2) of Table 2.3 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

Table 2.4: Interpretation of $\mathrm{E}(\nu)$ and $\mathrm{Var}(\nu)$

	$\mathrm{E}(\nu)$	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 $\mathrm{E}(\nu)$ 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.

with low insurance scores.

To help interpret and provide context for the KMS 95 percent confidence intervals for $\mathrm{E}(\nu_i)$ and $\mathrm{Var}(\nu_i)$, Table 2.4 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 2.4.3, 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 2.4 for the models in Cohen and Einav (2007) and Barseghyan et al. (2013) are

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

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 2.4 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 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 2.4 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 2.4 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 2.4 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 functions (though we note that this does not imply that the estimated choice set

[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 2.4 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 $\boldsymbol{\theta}_0 \in \Theta_I$.

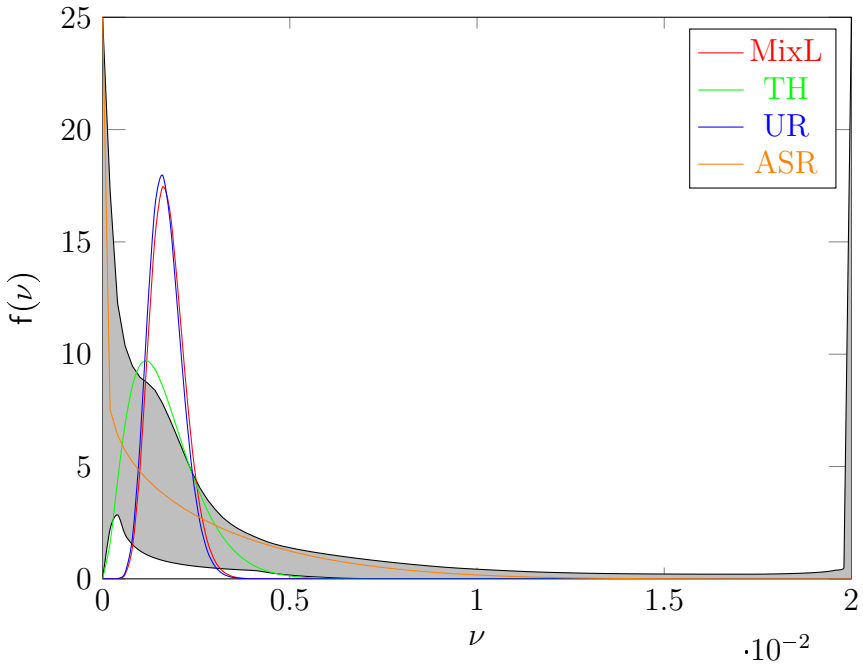


Figure 2.4: Confidence set for outer region of admissible probability density functions of ν.

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.

distribution obtained under the ASR model is not rejected).

2.5.3 Choice Set Size

Table 2.5 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

[22]By construction, due to the assumption that $\kappa = 3$ (Assumption 2.4.3(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

Table 2.5: 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.

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.

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,

upper bound on π_3 is one.

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]

2.5.4 Welfare

Lastly, Column (3) of Table 2.3 reports the upper bound of the KMS 95 percent confidence interval for the solution to problem (2.9). As explained in Section 2.3.3, 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 of coverage with a \$500 deductible (the modal choice) is \$217 (see Table 2.1). 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

[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).

higher/lower for households with low/high insurance scores ($60/$47) and considerably higher/lower for households with young/old principal drivers ($76/$36).

2.6 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:

$$\Pr(d_i = c|\mathbf{x}_i) = \sum_{G \subseteq \mathcal{D}} \Pr(c \in^* G|\mathbf{x}_i) \Pr(C_i = G|\mathbf{x}_i, c \in G), \tag{2.13}$$

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).

[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."

The two-stage characterization in equation (2.13) forms the basis of numerous models of discrete choice with unobserved heterogeneity in choice sets, including ours (as one can readily see from equation (2.3) 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 2.3.1). In order to point identify the model of preferences, which is represented by $\in^*$ in equation (2.13), the econometrician has to make assumptions—either explicitly or implicitly, sometimes arbitrary and often unverifiable—about the choice 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, macroeco-

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

[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.

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

[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."

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

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

[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.

Adams (2018), Barseghyan et al. (2019b), 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.

Barseghyan et al. (2019b) 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 en-

[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.

able 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 as-

sumption 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., 2019b; 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., 2019b) or evolve over time (Crawford et al., 2019), rely on exclusion restrictions or large support assumptions (Abaluck and Adams, 2018; Barseghyan et al., 2019b), 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.

CHAPTER 3

INSURANCE CHOICE WITH NON-MONETARY PLAN ATTRIBUTES: LIMITED CONSIDERATION IN MEDICARE PART D

3.1 Introduction

Health insurance markets in the United States are moving towards increased consumer choice. Many employers today offer their employees a choice of sponsored health insurance plans. The Balanced Budget Act of 1997 provided Medicare beneficiaries the opportunity to receive their health benefits through private insurance plans, and the Medicare Modernization Act of 2003 expanded those plans into what is today known as Medicare Advantage, or Part C. Since 2006 Medicare beneficiaries have the choice of prescription drug plans offered by private companies through Medicare Part D. Recently, following the Affordable Care Act of 2010, more individuals are choosing among private insurance plans through the expanded Medicaid program and online health exchanges. Few markets compare in economic magnitude to health care: in the United States, health care spending accounts for approximately 18% of GDP and continues to grow. The insurance products available and the corresponding choices individuals make in such markets have a large impact on their access to quality health care and overall well-being.[1] Public policy considerations surrounding health care and insurance are top of mind for many and are widely debated in contemporary politics. Efforts to improve the outcomes for individuals in health insurance markets must confront market inefficiencies, such as market

[1]Prescription drug insurance alone has been shown to improve health outcomes. Diebold (2016) and Semilla et al. (2015), document substantial improvements in drug adherence and mortality rates among beneficiaries enrolled in Medicare Part D.

power and asymmetric information. Depending on their nature, these inefficiencies may or may not require policy interventions in order to improve health and market outcomes. To this end, an understanding of the foundations of individual choice in health insurance markets is crucial to assessing the impact of any new policies, interventions, or modifications to market design.

Rationalizing health insurance choices is, however, notoriously difficult. Many choice patterns defy notions of optimality under economic models. It is not uncommon for individuals to select insurance plans that are strictly dominated by available alternatives.[2] In some settings, choices indicate preferences for attributes that do not conform to most economic models. The classic approach to insurance views contracts as lotteries with purely financial outcomes. Insurance appeals to risk averse individuals as a means to transfer wealth from good states of the world, in which they are not sick, to bad states of the world, where health needs are costly. In practice, this view of insurance is challenged by empirical patterns. Numerous studies of prescription drug coverage choice in Medicare have encountered such patterns: beneficiaries appear to overweight premiums relative to out-of-pocket costs and ascribe value to both non-monetary attributes and monetary attributes above and beyond their financial impact.[3] During early years of the program, the average beneficiary faced a choice from approximately 50 insurance plans.[4] In a market setting with such a large choice set of complex products, beneficiaries and policymakers alike have

[2]In a relatively straightforward comparison of employer-provided health insurance plans where plans differed in deductible and premiums, and thus only require a dollar comparison across plans, Bhargava et al. (2017) finds a substantial portion of individuals select plans that are strictly dominated regardless of preferences or health realizations. Handel (2013) documents substantial inertia in employer-provided health insurances leading to dominated choices, albeit with reduced adverse selection.

[3]See, for example, Abaluck and Gruber (2011), Ketcham et al. (2012), and Heiss et al. (2013).

[4]After adjustments to the market regulations, at present, on average beneficiaries face approximately 30 plans.

expressed concern that the choice environment is difficult to successfully navigate.[5]

In this paper, I propose an empirical model of demand for prescription drug plans where non-monetary plan attributes stochastically determine the composition of the set of plans that an individual considers, and monetary plan attributes determine the individual's expected utility over contracts in her consideration set.[6] This model reconciles the classic view of insurance contracts as lotteries with purely monetary outcomes with the empirical finding that choice among insurance plans is driven by their non-monetary attributes and financial attributes beyond their impacts on costs. This model of limited consideration, in which individuals are assumed to select their preferred plan from an unobserved subset of the feasible set, preserves the structural interpretation of insurance demand as arising from risk aversion, while providing a natural role for various plan attributes to shift choice frequencies in ways beyond the impact of those attributes on the utility derived from a plan. In what follows, "choice set" denotes the full available menu of plans, and "consideration set" refers to the subset of plans considered. I estimate the model using data from Medicare Part D allowing for unobserved heterogeneity in risk aversion and in consideration sets. Incorporating limited consideration into an expected utility model of insurance demand provides an avenue for the data to identify the elements of the choice environment that underpin limited consideration. The model determines the causes of limited consideration, such as the plan attributes, but does not presume a specific underlying behavioral model of consideration set

[5]See survey results in TheKaiser Family Foundation and the Harvard School of Public Health (2006), for example.

[6]I use the term "non-monetary" attributes throughout the paper in reference to both attributes that do not have an immediate monetary interpretation, as well as to the role of financial attributes above and beyond their impacts on costs.

formation.

Interest in the role of human cognition and assumptions regarding which feasible alternatives an agent considers when making a choice has a long history, including Tversky (1972) and Manski (1977). Models of limited choice sets have been a part of the literature on marketing for decades, as in Roberts and Lattin (1991) and Ben-Akiva and Boccara (1995). More recent developments in economic models, and specifically those in the framework of decision-making under risk, are described in Section 3.3 below. This paper leverages this history and recent results regarding consideration and risk preferences in Barseghyan et al. (2019b) to obtain point identification of a structural model of insurance choice alongside limited consideration. Moreover, the model is tractable to implement, even when the choice set is large. There are many potential underlying sources of limited consideration in the Medicare Part D market. Individuals may face constraints unobserved to researchers that result in the exclusion from consideration of certain plans deemed unfeasible. Many individuals face liquidity constraints and are unable to cover large unplanned expenses.[7] It is certainly imaginable that such a constrained individual might only consider plans with reduced or eliminated deductibles. Similarly, some beneficiaries live on a fixed income and a budget-constrained individual may only consider plans with monthly premiums below a reservation price. Market forces such as firm advertising or agent steering effects may lead beneficiaries to consider only plans offered by certain firms. Others may simply face cognitive or time limitations that manifest in a reduced number of plans considered at the time of enrollment. I remain agnostic about the behavioral mechanism behind limited consideration, and employ a consideration set formation model that allows for any of these

[7]See discussion in Durante and Chen (2019) within the section*Dealing with Unexpected Expenses*.

mechanisms to play a role.

A model of expected utility with limited consideration is well suited to explain plan choice patterns among Medicare Part D beneficiaries. Using a sample of beneficiaries living in the largest of the standalone prescription drug plan (PDP) regions, I recover estimates of risk preferences while allowing the probability a plan is considered to depend on the attributes highlighted in previous literature. Heterogeneity in consideration sets plays a crucial role in rationalizing plan choices. Beneficiaries in my sample face the choice of 46 plans, but over 90% of individuals consider no more than 5 plans. The probability a given plan is considered is driven by the identity of the insuring firm, the premium, the deductible, and the presence of supplemental coverage in the infamous "donut hole" (a phase of coverage in which beneficiaries pay 100% of drug costs). I estimate the highest premium plan is considered 10% as much as the lowest premium plan, all else equal. Similarly the highest deductible plan is considered 18% as often as a comparable zero deductible plan. In contrast, attributes that are not as easily observed by beneficiaries, such as the number of popular drugs covered, do not play a role in consideration. The consideration impacts of the firm and deductible alone are appreciable. Just three firms account for over 60% of considered plans, while the three smallest account for fewer than 0.5%. Although the majority of plans offered in the market include a deductible, nearly 75% of considered plans have no deductible. These patterns of consideration result in beneficiaries clustering on lower premium and zero deductible plans offered by a few popular firms that are not necessarily as well matched to their drug needs as other available but unconsidered plans.

In contrast to the previous Medicare Part D literature, I recover substantial

estimates of risk aversion in line with the literature that estimates risk aversion in field data.[8] My estimates more than double the mean risk aversion implied by a classic model of full consideration. My model highlights the sensitivity of risk preference estimation to the treatment of consideration. The material role of limited consideration, taken together with the distribution of risk aversion, translates into an important cost of limited consideration because beneficiaries frequently do not consider their best plans. Beneficiaries lose, on average, $226 in certainty equivalent terms, from considering a subset of plans that often does not include the plan best suited to their drug needs and risk preferences.

My estimates of risk preferences suggest a distribution of optimal choices that differs substantially from the empirical distribution. Estimates of plan consideration probabilities bridge the gap between these two distributions. The plans that are optimal for a large share of beneficiaries but are infrequently chosen are found to have relatively low consideration probabilities. Correspondingly, the most highly considered plans are those that are optimal for a relatively small share of beneficiaries and yet are often chosen. These estimates contribute to the primary source of the cost of limited consideration - by considering so few plans, individuals often do not evaluate plans that are best according to utility. Using my estimated structural model, I show that a counterfactual where certain consideration effects are removed leads to, holding all else equal, increases the size of consideration sets in the population and improves choice quality. The elimination of the firm effects, for example, more than triples the average consideration set size and increases the likelihood individuals consider their optimal plan.

[8]For example, Cohen and Einav (2007) finds among Israeli auto insurance customers a relatively low average risk aversion but a substantial fraction of customers exhibit very high risk aversion. Barseghyan et al. (2013) finds overall high levels of risk aversion among North American auto and home insurance customers.

This setting, in which economic theory suggests monetary attributes are the only utility-relevant plan features, but empirical patterns contradict that modeling assumption, previously created a dilemma for researchers. The model of limited consideration resolves some of the inconsistencies that have become commonplace in modeling insurance choices. Estimates and model implications are sensitive to the treatment of non-monetary attributes, and the usefulness of estimates of risk aversion without accounting for plan attributes is limited. This paper provides a tractable alternative to modeling insurance decisions that both preserves the role of risk preferences and guides policymakers towards how beneficiaries are navigating this complex choice environment. My results indicate that documented sub-optimal choice patterns are not a sign that the trend of increasing the role of consumer choice in health insurance is a lost cause. Accounting for limited consideration clarifies that consumer choices are not inexplicable, but rather reflect the navigation of a large, complex choice environment and the importance of certain easily ascertained features of the plans offered. Accordingly, the impact of adjustments to market regulations, plan design, or the manner in which plan information is presented to beneficiaries, will depend both on true risk preferences and the effect of such adjustments on consideration.

3.2 Institutional Background

Prior to the Medicare Modernization Act of 2003, Medicare provided hospital (Part A) and physician services (Part B) insurance coverage for elderly Americans and those with disabilities and certain serious illnesses. In 2006 prescription drug coverage was added to the program. Beneficiaries seeking prescrip-

tion drug coverage have the option of enrolling in a standalone prescription drug plan (PDP) through Medicare Part D or to bundle prescription coverage with the other health insurance through Medicare Part C (also known as Medicare Advantage). Any individual enrolled in either Parts A or B is eligible for coverage through Part D. Both Medicare Parts C and D are regulated by the Centers for Medicare and Medicaid Services (CMS) but provide beneficiaries a choice among plans offered by private insurance companies. To mitigate adverse selection, for every month an eligible beneficiary does not enroll in Part D, a penalty is accrued and applied as a perpetual surcharge upon eventual enrollment.[9] The penalty is the same regardless of which plan is ultimately chosen and is typically deducted directly out of social security benefits.

Participants in Part D select a plan for the following year between October 15th and December 7th during annual open enrollment. Those who do not qualify for low-income subsidies cannot change plans throughout the year.[10] The menu of available plans is determined based on which of the 34 CMS regions a beneficiary resides in.[11] Within each region, beneficiaries face a large set of plans to choose from, where the premiums are subsidized by the federal government and are fixed across individuals. As shown in 3.1, in 2010, regional choice sets varied from a minimum of 39 plans to a maximum of 54 plans.[12] Firms participating in a market can offer multiple plans and have some discretion over ways to differentiate their plans. All plans offered through the program must meet CMS requirements on minimum plan generosity, including covering at least 2

[9] If an eligible beneficiary receives prescription drug coverage that meets CMS standards through another channel, such as an employer program, this penalty is not amassed.

[10] Recently CMS has relaxed this rule slightly. Individuals are permitted to change plans throughout the year if they are moving into a plan CMS rates as 5-star in terms of quality.

[11] There are additional regions covering beneficiaries living in United States territories.

[12] In the data description below in Section 3.4, the plans listed here include only standalone PDPs, without an employer waiver, and exclude plans that were discontinued midyear due to CMS intervention.

drugs within 148 therapeutic categories, and virtually all drugs within certain crucial therapeutic classes.

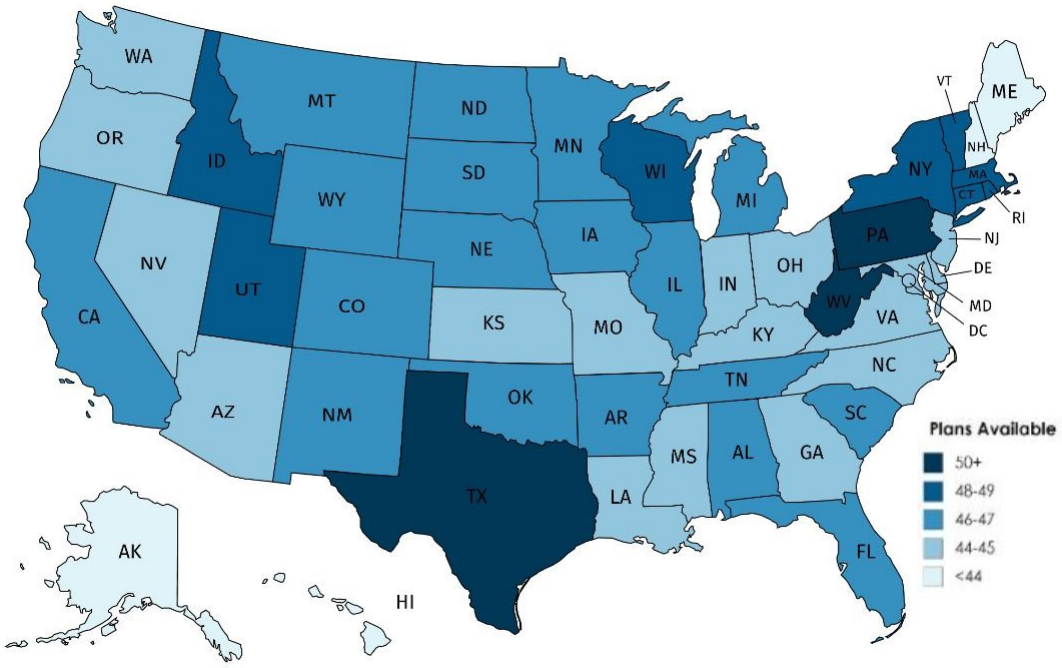


Figure 3.1: Counts of Plans Offered, 2010

Every year CMS releases cost-sharing standards for a base plan design. All plans in the program are required to be at least as generous actuarially as the standard plan. The standard plan divides beneficiary spending into four phases: the deductible, the initial coverage phase, the coverage gap (known colloquially as the "donut hole"), and the catastrophic coverage phase. Figure 3.2 provides a graphical representation of the 2010 standard plan. During the deductible phase, a beneficiary is responsible for 100% of drug costs. Once the deductible of $310 is reached, the plan's initial coverage begins, during which the plan covers 75% of drug costs and the beneficiary pays the remaining 25% out of pocket. Once the initial coverage limit of $2,830 is reached, a beneficiary enters the coverage gap where 100% of costs are borne by the beneficiary until an out-

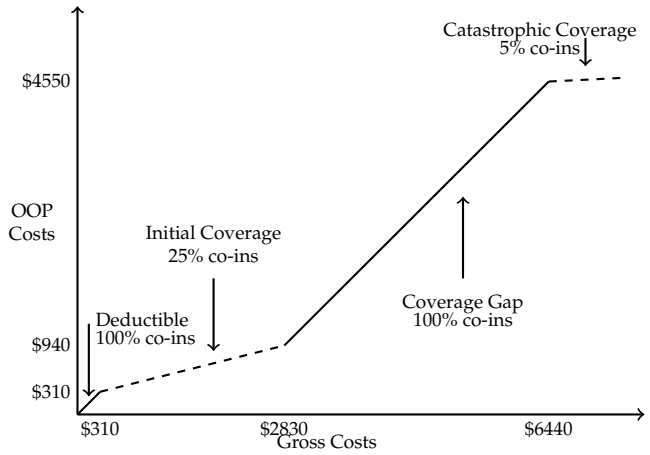


Figure 3.2: 2010 Standard Plan Design

of-pocket threshold of $4,550 is reached.[13] Any claims beyond the out-of-pocket threshold are treated as catastrophic and the beneficiary pays the maximum of a $6.30 copay or a 5% coinsurance.[14]

While the market is highly regulated, firms have the ability to differentiate the plans they offer. Market regulations during most of the program's existence limit the number of plans a firm can offer in a given market and require a meaningful level of distinction between plans offered by the same firm to avoid confusion from seemingly redundant plans. There are multiple ways a firm can differentiate the plans they offer from one another and from those offered by other firms in a region. Insurers have wide discretion over the plan formulary, which lists all drugs covered under a plan and how generously they are covered by classifying each included drug into a tier (lower tiers correspond to lower cost drugs). Firms can also adjust the cost-sharing structure of a plan, with many choosing to offer plans with a reduced or fully eliminated deductible. Under

[13] As part of the Affordable Care Act, the coverage gap was mandated to be phased out, absorbed into the initial coverage phase, over 2011-2020. It was fully eliminated a year ahead of schedule in 2019.

[14] For branded drugs the copay is $6.30 and for generic drugs it is $2.50.

such a design, claims are processed according to the initial coverage structure from the first dollar.

Although with some exposition here the government designed plan appears relatively comprehensible, this simple plan description belies some of the further complexities of the products beneficiaries face. Consider, for example the deductible of a given plan. In the standard plan, this is described simply as a dollar amount up through which costs are borne fully by the beneficiary. Each year CMS determines a maximum deductible allowed under the standard plan, but firms can offer reduced or zero deductible plans (and in the cases where a firm is offering multiple plans, they can use differing deductible amounts to differentiate those plans). Many of the plans with a deductible exempt low cost drugs (categorized as tier 1 or 2) from the deductible, and rather process them under the standard cost-sharing used during the initial coverage phase. To understand the impact of the deductible on out-of-pocket costs, a beneficiary must have an understanding of the timing of their claims, as well as which tier classification has been assigned to their needed drugs under different plans, since plans have discretion over this classification. This complexity requires that a beneficiary evaluating two plans that at first blush appear to differ only in the deductible phase, may still be facing a rather complex comparison. To ease this process, CMS encourages beneficiaries to use the online PlanFinder tool, where beneficiaries can enter their zip code, expected drug needs, and pharmacy preferences to receive personalized estimates of out-of-pocket costs under each available plan.

Despite these complexities, Part D has been, on the whole, lauded as a success. Studies, including Diebold (2016) and Semilla et al. (2015), have found

substantial improvements in prescription drug adherence and mortality rates among beneficiaries enrolled in the program. The program is widely used and popular among beneficiaries, with 43 million beneficiaries enrolled in 2018. From the start, however, there has been concern that the plan choice environment is too complex, especially for a more senior population. Beneficiaries themselves expressed interest in a reduced choice set in order to alleviate the difficulty in choosing a plan.[15] The number of plans offered has decreased from the initial years of the program with the average beneficiary now facing a set of approximately 30 plans. The program remains popular with the majority of Medicare beneficiaries receiving prescription drug coverage through it.

3.3 Literature Review

This paper shares a core motivation with previous studies on Medicare Part D plan choice: to understand and evaluate plan choices according to economic models of decision making. Well known studies include Heiss et al. (2013) and Abaluck and Gruber (2011), as well as the exchange resulting from the latter in Ketcham et al. (2016) and Abaluck and Gruber (2016a).[16] This paper differs methodologically from such prior studies. Abaluck and Gruber (2011) evaluates initial plan choices in 2006 using data from a switch agent. Abaluck and Gruber (2011) estimates a conditional logit as a linear approximation of a CARA expected utility model with plan attributes included additively. Although in-

[15] For example, The Kaiser Family Foundation and the Harvard School of Public Health (2006) notes that in the first year of the program 73% of seniors found the program too complicated, as did 91% of pharmacists and 92% of doctors surveyed. 60% of seniors agreed that Medicare should select a small number of good plans to help seniors have an easier time choosing.

[16] Additionally important early studies include Heiss et al. (2010), Lucarelli et al. (2012), Kling et al. (2012), and Ketcham et al. (2012).

corporating plan attributes into the utility framework improves the explanatory power of the model, the resulting estimates are challenging to interpret in the classic insurance model, for reasons I now discuss. The resulting coefficients of those attributes are compared to those of premiums or out-of-pocket costs to monetize the attribute and assess an approximate willingness to pay. There are reasons to assume a beneficiary may ascribe a "cost" to certain attributes - for example, paying a deductible may cause disutility due to liquidity constraints that make a large single payment particularly challenging. The modeling technique of adding the deductible as a term in the utility specification, however, suggests a constant utility "cost" of the attribute across all possible health realizations, similar to the premium.[17] If the utility relevance of the deductible is meant to capture a burden or hassle cost of the attribute, in some contexts it is undesirable to model that cost as equal in the state of the world in which the beneficiary is not sick and does not incur the deductible and the state of the world in which she is sick and pays the deductible. Inclusion of attributes such as the deductible can offer insight into which plan features relate to choice probabilities, but with the existing modeling approach this is at the expense of the economic interpretation. As I explain in Section 3.7.4, my proposed modeling approach resolves this tension.

Using a conditional logit model as described above, results indicate that beneficiaries are selecting plans in a manner considered inconsistent with the rational behavior of an expected utility maximizer who evaluates plans based on their monetary features. Specifically, beneficiaries overweight premiums relative to out-of-pocket costs, place little to no value on a plan's risk reduction features, and value financial aspects of plans, such as deductible and gap cov-

[17] See Handel (2013) and Handel and Kolstad (2015) for a discussion of this topic.

erage, beyond the impact of such attributes on expected costs. Drawing similar overarching conclusions, Heiss et al. (2013) estimates a multinomial logit model to approximate a CARA expected utility model including the theoretically relevant cost variables. Such a model without plan attributes poorly describes the choice patterns of beneficiaries in Medicare administrative data. Moreover, the implied risk preferences are surprisingly unstable over time, with one year of modest risk aversion and one year of substantial risk preference.[18] In both of these studies, a logit is used as a linear approximation to the CARA expected utility function.[19]

Given the complexity of prescription drug plans, the large number of plans available, and the advanced age of beneficiaries, it is hardly surprising that individuals would fail to behave in a manner fully consistent with standard economic models. The literature on menu complexity and heuristic shortcuts in insurance also shares motivational elements with this paper. In the Part D market, Ketcham et al. (2015) finds evidence that it is not the size of the choice set alone that drives choices inconsistencies. The quality of choices is estimated to improve with larger choice sets due to increased switching, with the exception of the cases where additional plans are relatively more expensive. In other health insurance markets, there is evidence that consumers use heuristic shortcuts to limit the choice set before choosing plans (Ericson and Starc (2012)), as well as that choices improve when products are standardized and the choice set becomes less complex (Ericson and Starc (2016)).

[18]Similarly to Abaluck and Gruber (2011), the coefficient of risk aversion is estimated based on a ratio of the estimated coefficients on variance of costs and mean costs. In 2008, Heiss et al. (2013) estimate a positive coefficient on variance, implying riskier plans correspond to higher choice probabilities.

[19]For a derivation of such an estimating model from the CARA expected utility framework, see Abaluck and Gruber (2011).

Inconsistencies with model implications can also be suggestive of model misspecification. Ketcham et al. (2016) implement a very general test of rationality, using General Axiom of Revealed Preference (GARP) arguments to determine if plan choices are consistent with any utility specification. Although focused predominantly on highlighting that the majority of plan choices are consistent with some utility function, and thus evidence of widespread suboptimality of plan choices is potentially indicative of model misspecification, the fact remains that even under such a general framework a sizable fraction of initial plan choices remain inconsistent with utility maximization.[20] Many of the studies on Part D plan choice have differed in model, data, and measures of choice quality, but there is an empirical consensus that seniors are leaving money on the table.[21] Such deviations from rationality are not unique to prescription drug insurance choices. Bhargava et al. (2017) describes a case of employer offered health insurance plans in which a substantial portion of individuals select insurance plans that are strictly dominated by available alternatives, and in such an unambiguous manner that basic arithmetic would highlight that dominance.[22] In the market for auto collision insurance Barseghyan et al. (2016), Barseghyan et al. (2019b) and Barseghyan et al. (2019a) document a substantial fraction of individuals selecting a policy that is dominated by other available plans, regardless of risk preferences.

A commonly suggested and intuitive explanation for the prevalence of what

[20] In the main version of this test, 21% of choices over 2006-2010 were inconsistent with utility maximization. The further relaxed consistency test found 14% of those choices are not rationalizable.

[21] See in addition Ketcham et al. (2015), Kesternich et al. (2013), Kling et al. (2012), among others.

[22] The setting included pairs of plans differing only in the deductible, but the additional premium charged for the lower deductible plans exceeded the amount by which the deductible was reduced, guaranteeing larger costs under all realized health scenarios.

economists deem suboptimal choices is limited consideration.[23] In a model of limited consideration, individuals are assumed to select a plan (or product, more generally), from a considered subset of the feasible set. Choices, therefore, do not reveal preference over the entire choice set, but rather only over the considered set. Previous studies, including Abaluck and Gruber (2011) and Abaluck and Gruber (2016b), have mentioned limited consideration as a possible explanation for the role of plan attributes in choices.[24] A well-studied and generally accepted form of limited consideration in the Medicare Part D market is inertia.[25] Fundamentally, inertia is an type of limited consideration in which the agent considers only their existing plan or no plans at all. Ho et al. (2017), studying Medicare Part D choices over time, documents the role of inertia and the way in which certain shocks - most notably in premium, a highly visible plan attribute - can break beneficiaries from their inertia. Abaluck and Gruber (2016b) also study Medicare Part D choices over time, and documents a role for inertia and finds little evidence of learning or improved performance of beneficiaries as they gain more experience in the market over time. Polyakova (2016) explores the interaction of inertia, adverse selection, and market regulations and finds inertia and switching costs contribute to the sustainment of an adversely-selected equilibrium in Medicare Part D.

Additional explanations for observed choice patterns in the market have recently been explored in the literature. Keane et al. (2019) and Ketcham et al. (2019), for example, propose an alternative approach in which Part D choices

[23]Models of limited consideration and limited attention have a long history in economics, including Simon (1959).

[24]Both studies posit that the importance of firm fixed effects in matching choice patterns may suggest individuals are rationally using trusted firms as a heuristic shortcut when unable or unwilling to make the time-consuming or difficult financial comparison across all plans.

[25]Inertia is the well-documented pattern of behavior in insurance markets whereby individuals passively remain in existing plans at the time of a renewal rather than actively select from the set of available plans.

are assumed to be made with varying degrees of consumer informativeness. In the former a a mixture-of-experts model is used to model plan choices as probabilistically revealing of preferences. The latter uses survey data as a signal of whether consumers are informed and assess the welfare implications of various market interventions assuming the observed choices of informed individuals proxy for the preferences of uninformed individuals. In contrast, Brown and Jeon (2019) build on the work of Matějka and McKay (2015) and Fosgerau et al. (2017) and propose a model of rational inattention whereby beneficiaries for whom the choice stakes are high, those with high variance of out-of-pocket costs across available plans, acquire more information about the plans before enrollment.

Beyond the framework of health insurance, this paper builds on the methodology of limited consideration in discrete choice models. The alternative specific consideration model used below has been developed and shown to be nonparametrically identified under certain conditions in Barseghyan et al. (2019b).[26,27] This paper highlights a major appeal of such models to empirical applications. The introduction of consideration sets provides a natural role for non-monetary plan attributes in a model of insurance choice - the probability a given plan is considered can be modeled as a function of its attributes. This is the key distinction of this paper compared to previous studies. In such an insurance model, there is a distinction between what enters expected utility and reflects the uncertainty of the environment and the plan attributes that enter consideration and hence do not depend on the state of the world. In cases where plan attributes

[26]The model in Barseghyan et al. (2019b) expands on the work of Manski (1977) and Manzini and Mariotti (2014).

[27]A conceptually similar, but econometrically different and parametrically specified, model was used in Goeree (2008) to estimate demand for computers when advertising plays a role in consideration set formation.

impact choice beyond their financial impact on utility-relevant monetary costs, consideration sets provide a theoretically sound avenue to relate attributes to choice. A limited consideration model can marry the theoretical underpinnings of expected utility with the empirical reality present in this market. These techniques offer a tractable modeling alternative to standard methods, even in the presence of such a large feasible choice set.

3.4 Data

The primary data source in this study is administrative data from CMS. These data include information for a 5% random sample of 2010 Medicare beneficiaries. The relevant enrollee data include information on basic demographics, plan choice, and the full set of drug claims filled under the beneficiaries' plans. These beneficiary and claims data are paired with plan information, linking premiums and plan coverage structures for all plans available to each beneficiary. Additional information on the formularies and drug prices negotiated for each plan is included in public use files released for purchase by CMS, as well as the restricted access version of the formulary file for 2010.[28] Official firm names listed in the restricted files are matched to the common company names beneficiaries would see at the time of plan choice, checked against a crosswalk published by Ketcham et al. (2016).

[28] Although the restricted version of the formulary file is available, surprisingly, information on negotiated base prices for drugs is only included in the public use versions of the data.

3.4.1 Analysis Sample

The aim of this study requires restricting the sample to beneficiaries selecting standalone prescription drug plans (PDPs), excluding those who forgo prescription drug coverage, those with coverage outside of Medicare and those who opt instead for Medicare Advantage (Part C) plans.[29] Additionally, I exclude from the sample all individuals receiving a low-income subsidy. The enrollment, pricing, and choice environments for those individuals differ substantially from the standard Medicare population, and I lack relevant data on payments. I also exclude individuals who have more than one Medicare drug plan over the course of the year, are dual eligible for Medicaid, or drop their coverage mid-year for any reason other than death. As a final general sample restriction, I exclude individuals who either currently have or initially enrolled in Medicare due to end-stage renal disease, as their health needs differ quite dramatically from the overall Medicare population.

Every year during open enrollment, beneficiaries select a plan for the entirety of the following year.[30] If a beneficiary's existing plan remains available, they default into the same plan without an active choice. A concern in any choice environment of this sort is the distinction between the role of preferences and the role of inertia in observed choices. The role of inertia is left for future research, and for this paper, I abstract from this complication by restricting attention to "active choices." Active choices include the enrollment decisions of those joining Medicare Part D upon eligibility, as no default option is available.

[29] My data include information on whether individuals receive outside coverage but lack any specific information on offerings, pricing, and claims for such cases.

[30] Beneficiaries receiving low-income subsidies are permitted to change plans monthly. The beneficiaries within the present sample, however, cannot change plans until open enrollment, where they can select a different plan for the subsequent year.

Additionally, I include individuals that are first choosing a Part D plan but for a few common reasons, are not making that choice at the time of eligibility. This includes those that either retained employer drug coverage (through the form of a Retiree Drug Subsidy plan) for a period of time after entering Medicare, initially retained other creditable prescription coverage while Medicare enrolled, or went for a period of time without any drug coverage. If these individuals joined Part D during 2010, their choices are included in my sample.[31] The final group included as active choosers are those who are actively switching plans from the previous year.[32]

The set of plans available to beneficiaries is determined by the region of residence. This study focuses on active choices among residents of California (Region 32), the largest of the PDP regions. Table 3.1 presents summary statistics of the full 2010 active choosers sample and the 2010 California subsample. Californians in 2010 could choose from 46 plans, offered by 19 different firms. On the whole, the California beneficiaries are similar along characteristics to their national counterparts, but differ along choice patterns in specific dimensions. On average, California beneficiaries are less white, slightly younger, and, correspondingly, file fewer claims. In a similar fashion to the national average, these beneficiaries are largely choosing plans offered by the most popular firms. These individuals, however, enrolled in zero deductible plans in much larger numbers than the average active chooser in the US. The majority of the 46 plans available to the California beneficiaries include a deductible, yet 69% of the sam-

[31] This latter group does face the above described penalty upon enrollment, but the resulting surcharge is constant across all plans and is typically charged directly out of Social Security payments. As such, I abstract away from the role of the penalty on the choice of plan.

[32] Without the Part D Plan Election Type Beneficiary Summary File, it is difficult to determine if a plan change is an active choice or a passive transition upon termination of the existing plan. Conservatively, I include as active choosers those switching plan types, for example from an HMO to a PDP, from 2009 to 2010, as well as those who select a 2010 plan offered by a different firm than their 2009 plan.

ple enrolled in a plan without a deductible.

Table 3.1: Summary Statistics: Active Choosers

	U.S.	CA
Sample Size	69,278	4,412
2010 Months Covered	9.9	9.3
Age	71.0	69.8
Female	.584	.566
White	.936	.889
Black	.041	.018
Hispanic	.004	.013
Asian	.007	.035
Monthly Claims	2.5	2.3
Days Supply	43.2	42.3
Average Total OOP	$625	$635
Average Total Gross Costs	$1,727	$1,639
Number Plans Offered	46.6	46
Avg Deductible	$97.18	$64.00
Zero Deductible	.539	.690
Avg Monthly Premium	$35.56	$38.12
Top 1 Most Popular Firm	.317	.372
Top 2 Most Popular Firms	.507	.505
Top 3 Most Popular Firms	.613	.632
Min Premium within Firm	.449	.306
Min Deductible within Firm	.578	.722
Min Premium or Deductible within Firm	.927	.985

Note: Statistics computed over "active choosers" in the 2010 sample based on description above. All statistics reflect unweighted averages.

The plans offered in the California market exhibit substantial variation in attributes previously documented as choice-relevant in the literature. The 46 available plans were provided by 19 different insurance firms; 16 plans included

the maximum deductible, 11 included a reduced deductible, and 19 did not included a deductible; 20% of the plans offered some form of coverage in the donut hole. Of the 100 most popular drugs by sale among beneficiaries, the plans in California covered between 71 and all of them, with an average of approximately 91 drugs covered. The plans offered varied in average cost-share in the initial coverage phase from 33% to 58%. Among this large and varied choice sets, beneficiary choices were fairly concentrated. Only 16 of the 46 plans garnered enrollment in excess of 1% of the sample.

3.4.2 Distribution of Expected Out-of-Pocket Costs

Although the CMS data is rich, it only contains claims and spending information for beneficiaries under their chosen plans. To estimate a model of plan choice, I require the counterfactual costs beneficiaries would face under the set of alternatives available to them, as well as a measure of the variance of out-of-pocket (OOP) costs. To estimate these counterfactual costs, I construct a plan calculator that takes in any specified set of claims for an individual and computes the out-of-pocket expenses that the specified sequence of claim events generates under every plan available. Consider an individual who fills a number of prescriptions each month. Under each available plan's formulary, each of those drugs claimed is classified by tier that determines the cost-sharing structure used, whereby cheaper drugs are assigned a lower tier than more costly drugs. Additionally, within each plan a different base price of the drug has been negotiated to which the plan's cost sharing structure is applied. The calculator procedure involves determining the tier each plan assigns a drug and calculating the out of pocket costs for each claim accounting for the cumulative

claim and corresponding coverage phase.[33]

The purpose of the calculator is to quantify counterfactual spending distributions under the set of available plans with the understanding that a rational beneficiary would compare plans in terms of the out-of-pocket costs in each plan. It is not obvious what sequence of claims an individual anticipates at the time of plan choice. Some have assumed that beneficiaries have "perfect foresight", and assume that at the time of plan choice, beneficiaries compare the expected out-of-pocket costs of the drugs they would come to claim during the year of coverage.[34] Alternatively, some studies have assumed a myopic approach, assuming that beneficiaries base their expectations on their previous year drug claims when data is available or current drug needs.[35] It is also possible to take a "rational expectations" approach and assume individuals predict their drug needs will be realized from a distribution of costs under each plan based upon the realized costs of a set of "similar" individuals. For expected mean expenditures, the results in Section 3.7 adopt a perfect foresight assumption. The robustness analysis in Appendix B.3 presents results under a myopic approach, projecting the first month of claims experience in 2010 for the remainder of a beneficiary's time in the plan (note the popular and CMS-promoted

[33]There are multiple numeric codes used to identify drug by molecule, formulation, and strength. These numeric systems do not, however, identify drugs uniquely. The claims data identifies drugs by National Drug Codes (NDCs), as well as a CMS created number referred to as the Formulary RX ID. The public use formulary data identify drugs by NDCs and RXCUIs. However, multiple NDCs can be used for the same drug. As such, NDCs are considered the same if they are linked through Formulary RX IDs. For example, consider a drug denoted as NDC_1 and $FRXID_1$. If NDC_1 is also listed as corresponding to $FRXID_2$, and $FRXID_2$ is elsewhere linked to NDC_2, I consider NDC_1 and NDC_2 the same drug. For each claim passed through the calculator, I apply the lowest tier and base price of any linked NDC, allowing for some potential, albeit minor, substitution.

[34]For example, ex post claims are used as the anticipated mean out-of-pocket costs under plans in Abaluck and Gruber (2011) and as one of two alternative models in Abaluck and Gruber (2016b).

[35]See, for example, Kesternich et al. (2013), Heiss et al. (2013), and Abaluck and Gruber (2016b).

online tool to help with plan choice, PlanFinder, uses this approach).

Any measure of higher order moments of the distribution of expected costs requires the latter approach of binning similar individuals. To estimate a distribution of out-of-pocket costs an individual in the analysis sample expects, beneficiaries are grouped into bins of "similar" individuals based on average monthly number of claims and average monthly gross cost of claims.[36] Details of this procedure are outlined in Appendix B.1. A random sample is drawn from each bin and their claims are passed through the plan cost calculator to estimate a distribution of costs under each plan. The higher order moments of the cost distribution that enter an individual's utility function are computed from this sample distribution of similar individuals.

3.5 Reduced Form Evidence of Limited Consideration

In previous studies of Part D enrollment, even though data and models may differ, there is evidence that seniors are selecting drug plans that are more expensive for their drug needs than available alternatives. It is challenging in empirical settings to distinguish between preferences and consumer "mistakes." To motivate the model described in Section 3.6, I conduct a reduced form analysis to show that the patterns of choice inconsistencies documented in previous studies manifest in my sample. Table 3.2 presents statistics on plan choices among the California sample. The top panel performs a GARP-style test of rationalizability following Ketcham et al. (2016). I compute the share of individuals

[36] In cases where claims correspond to multiple months' supply of drugs, we treat it as multiple claims. For example, a claim for a 90 day supply of a drug is treated as 3 claims in this exercise.

selecting plans on the mean of OOP expenditure frontier, the mean and variance of OOP expenditure frontier, and the mean-variance-firm frontier.[37] By focusing on dominance, these measures test whether choices are consistent with *some* utility function rather than a certain specification. In the classic insurance framework, the monetary cost variables (and potentially higher order moments of the cost distribution) are considered utility relevant. These statistics display that the individuals in this sample are not selecting optimally according to the monetary plan attributes economic theory suggests are relevant.[38] Fewer than 17% of beneficiaries select the lowest cost plan for their realized drug needs. Using a relaxed measure, I find approximately 24% of beneficiaries select a plan within 5% of their minimum cost plan, and around 30% within 10% of the minimum cost plan.

It is consistent with standard insurance theory for a risk averse individual to pay more for less variance in expenditures. Evaluating plan choices on the mean-variance frontier implies choices are dominated only if there is another plan available that is at least as good in terms of mean and variance of expenditures and strictly better in at least one of those measures. Rationalizability of observed choices improves by this measure but the majority of plan choices remain dominated. This means that, on the whole, the foregone savings are not offset by a reduction in risk. A major boost to explanatory power comes from incorporating preferences for a specific firm. In this formulation 89% of plan

[37]For the sake of comparison to previous studies, I use a perfect foresight model of expected costs whereby the realized 2010 claims for each individual is priced through each available plan. If instead the mean of a random sample of "similar" individuals is used or the first month of drug needs is projected forward, the patterns remain.

[38]Overall the rationalizability of plan choices is higher in this sample than previous studies. This may be due to improved choice performance in 2010 relative to earlier years, or sampling criteria. In contrast to earlier studies, my sample includes fairly young beneficiaries, partial-year enrollees, and active switchers. It is possible those groups choose slightly better than the average beneficiary in the first year of the program.

Table 3.2: Choice Rationalizability and Clustering on Certain Attributes

	% of Sample
Rationalizability Test	
Mean Cost Frontier	16.6
Mean-Variance Frontier	43.0
Mean-Variance-Firm Frontier	89.0
Attribute Choice Patterns	
Zero Deductible	69.0
Market's Lowest Premium	9.8
Min Premium within Firm	30.6
Min Deductible within Firm	72.2
Min Deductible or Premium within Firm	98.5
Gap Coverage	5.6
Top 1 Most Popular Firm	37.2
Top 2 Most Popular Firms	50.5
Top 3 Most Popular Firms	63.2

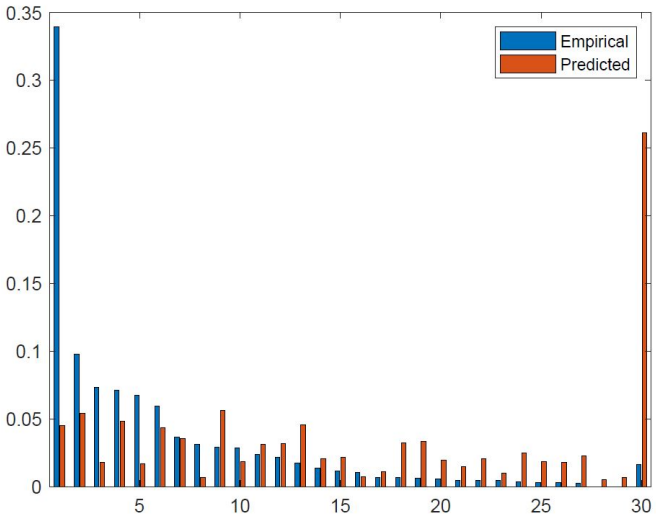
Notes: Mean cost computed based on assumption of perfect foresight of drug claims and monthly premiums prorated for total months of 2010 enrollment. Variance estimated from a distribution of 100 randomly sampled "similar" individuals as described in Appendix B.1. Most popular firms reflect the firms with the largest enrollment shares among the analysis sample.

choices are consistent with utility maximization of *some* utility function. This test of rationality designates a choice as a mistake if a beneficiary selects a plan that is dominated in the mean-variance space by another plan offered by the *same* firm. In this market, each firm offers typically 1-3 plans within a region, leaving little room for a dominating plan. And yet, 11% of this sample selects such a dominated plan.

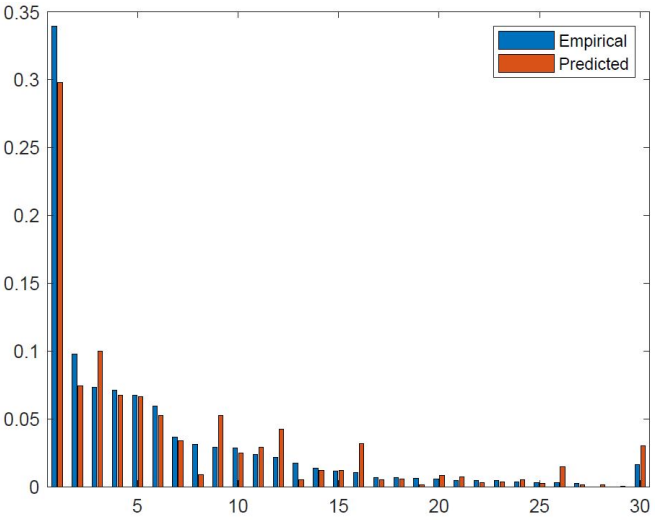
Another pattern that emerges in the lower panel of Table 3.2 is the prevalence of certain attributes among chosen plans. The majority of California beneficiaries select a plan without a deductible, even though, as described in Section 3.4.1, the majority of plans offered in California include a deductible. Beneficiaries are on the whole selecting plans with low deductibles, low premiums, and offered by one of the three most popular firms.[39] As a statistical test of explanatory relevance, Table 3.3 presents results of a simple logit regression. Column (1) includes in the regression the monetary variables describing the distribution of costs included in a standard model of insurance demand. Column (2) includes plan attributes and firm fixed effects. The Pseudo R^2 of the regression in Column (2) is approximately three times that of Column (1). Figure 3.3 graphically contrasts the explanatory power of these logit regressions by plotting the implied choice probabilities under each set of estimates. The right-most bar is a composite plan aggregating all 17 plans in which between 1 and 10 individuals in the sample enrolled. The improvement in fit with the additional attributes is visually obvious.

With the exception of the measure of average cost-share, the additional plan attributes in Column (2) are highly significant regressors. Some of these attributes - firm dummies, the count of top 100 drugs covered, and the average

[39]Generally, reduced deductibles come at the expense of a higher premium.



(a) Excluding Non-Cost Attributes



(b) Including Non-Cost Attributes

Figure 3.3: Logit Implied Choice Distribution

Notes: Panel (a) plots the model implied choice probabilities from Column (1) of Table 3.3 in red. Panel (b) plots the model implied choice probabilities from Column (2) in red. In both figures, the blue bars correspond to the empirical choice shares. Plans are ordered from the plan with the largest enrollment share on the left to the plans with zero enrollment. The rightmost plan corresponds to a composite plan of the 17 plans in which between 1 and 10 individuals enrolled.

Table 3.3: Conditional Logit Estimates: Impact of Non-Monetary Attributes

	(1)	(2)
Premium + $\mathbb{E}$OOP	-0.507***	-0.381***
(hundreds)	(.009)	(.011)
Variance	-0.022***	-0.007***
	(.002)	(.002)
Deductible	–	-0.634***
(hundreds)		(.026)
Gap	–	-0.767***
		(.076)
Top100 Drugs	–	-0.071***
		(.006)
Avg CS	–	-0.231
		(.611)
Firm Dummies	No	Yes
Pseudo R^2	0.130	0.353

Notes: Standard errors are in parentheses. Variance denotes the variance of $\mathbb{E}$OOP measured in hundreds of dollars. *** Significant at 1% level.

cost-share - are non-monetary attributes. There is no immediate way to compare these variables to those related to the costs of each plan. The deductible, while financial in nature, is also not directly related to costs. Insofar as the deductible of a plan impacts the costs of prescription drugs under each plan, it is already accounted for in the expected out-of-pocket cost term. The coefficients on the plan attributes above reflect a relationship between the attributes and

plan choice above and beyond their impacts on costs. In specifications where the coefficient on premium and expected out-of-pocket costs are permitted to differ, the estimates suggest an over-weighting of premiums relative to out-of-pocket costs. These estimates only reflect correlation but are informative for a structural model. The reduced form regressions in Table 3.3 indicate a model that includes only monetary attributes cannot rationalize observed choices as well as a model that accounts for additional non-monetary plan attributes.

There are many possible mechanisms through which these non-monetary attributes can affect plan choice. Some may find the large menu of plans burdensome and employ heuristics to reduce the choice set to a more manageable size. Other beneficiaries may act on their uncertainty over future needs by only considering plans with more generous coverage. Others, still, may have liquidity constraints and only consider plans with a reduced or eliminated deductible. Premiums may receive additional weight over expected out-of-pocket costs due to budget constraints. The model and procedure below do not require the researcher to take a stance on how exactly beneficiaries are paying attention to plan attributes. This agnostic approach is focused on flexibly approximating this process in order to learn what beneficiaries appear to be paying attention to when they make initial Part D plan choices and leaves to future work more precise exploration of the details underlying the consideration set formation process in this market.

3.6 Model of Plan Choice

3.6.1 Utility Specification

My model of plan choice maintains the expected utility framework standard in the literature of decision-making under uncertainty. Individuals are assumed to have utility over wealth and face a distribution of financial losses. In this empirical setting, each beneficiary has uncertain drug needs during the year and the coverage and cost structure of each available plan translates those drug needs into out-of-pocket costs. Denoting costs by C and initial wealth by W, the realized wealth of an individual is given by $W - C$. I assume individuals are risk averse, and their utility is governed by a coefficient of risk aversion, ν, assumed to be constant across values of wealth. This emits a utility model of constant absolute risk aversion (CARA) of the form:

$$U(C) = -\frac{1}{\nu}\exp(-\nu(W - C)),$$

Plans differ in whether and how generously drugs are covered and how cost-sharing is determined. These uncertain drug needs therefore correspond to different distributions of out-of-pocket costs under each plan. I denote the random variable of out-of-pocket costs individual i incurs under any plan j as $C_{ij} \sim F_{C_{ij}}$. Additionally, risk aversion is assumed to be heterogeneous across agents with $\nu_i \sim F_\nu$. The utility of individual i from choosing plan j is given by

$$U_{ij} = -\frac{1}{\nu_i}\exp(-\nu_i W_i)\exp(\nu_i C_{ij})$$

Under the assumption of expected utility, agents are assumed to take into account the distribution of financial losses they face and take an expectation of utility under each available plan. Conditional on a beneficiary's coefficient of absolute risk aversion, expected utility takes the form

$$EU_{ij} = -\frac{1}{\nu_i}\exp(-\nu_i W_i)\mathbb{E}\left(\exp(\nu_i C_{ij})\right)$$

Note that for a fixed value of ν_i, $\mathbb{E}\left(\exp(\nu_i C_{ij})\right)$ is the moment generating function of the random variable C_{ij}. Similarly to elsewhere in the literature, out-of-pocket costs are assumed to be Normally distributed, $C_{ij} \sim N\left(\hat{\mu}_{ij}, \sigma^2_{ij}\right)$, where $\hat{\mu}_{ij} = p_j + \mu_{ij}$ is the mean out-of-pocket expenditures of individual i under plan j, shifted by the person-invariant premium for plan j.[40] The cost parameters μ_{ij} and σ^2_{ij} are computed outside of the model as described in Section 3.4.2 and Appendix B.1. Substituting for the moment generating function, expected utility can be written as a function of the mean and variance of out-of-pocket costs.

$$EU_{ij} = -\frac{1}{\nu_i}\exp(-\nu_i W_i)\exp(\nu_i \hat{\mu}_{ij} + \frac{1}{2}\nu_i^2\sigma^2_{ij})$$

Although utility values depends on unobserved individual wealth, relative utility and the ordinality of plan utility are not impacted by the positive multiplicative term $\frac{1}{\nu_i}\exp(-\nu_i W_i)$. This value can be divided away from all utility levels and utility rankings remain unchanged. Therefore, for estimation purposes, a simpler form of expected utility suffices.

$$EU_{ij} = -\exp(\nu_i \hat{\mu}_{ij} + \frac{1}{2}\nu_i^2\sigma^2_{ij}) \tag{3.1}$$

[40] See Abaluck and Gruber (2011).

3.6.2 Choice Sets and Limited Consideration

Motivated by empirical findings that numerous plan attributes affect individuals' choices beyond the financial impact of those attributes on drug costs, the point of departure from a standard expected utility models is in consideration sets and the role plan of attributes in consideration. Rather than incorporating a random error into utility, stochastic choice, conditional on preferences, arises through the formation of the consideration set. Moreover, it is through the consideration set that plan attributes impact choice. As described in Section 3.3, previous studies have found evidence of plan attributes determining plan choice in ways beyond the experienced financial impacts of those attributes, but have struggled with a rational utility explanation of such a role. This model posits that these important, but not directly utility-relevant, variables impact choice by determining the composition of the consideration set a beneficiary evaluates when selecting a plan.

A model of limited consideration relaxes the assumption in standard discrete choice models that a chosen plan is revealed preferred to all available plans. Beneficiaries are assumed, rather, to select an unobserved subset of the feasible choice set to actively consider and compare and select a plan from that subset. I model an alternative specific consideration probability model, similar to that found in Barseghyan et al. (2019b). Under the assumption of limited consideration, an observed choice of plan j^* by individual i implies 2 things: 1) plan j^* was in individual i's consideration set, and 2) among all of the plans considered, j^* was preferred.

I denote beneficiary i's choice of plan j^* by $y_{ij^*} = 1$, and an individual's consideration set by M_i, which is a subset of the entire feasible set of plans $\mathcal{M}$.

The probability individual i chooses plan j^*, suppressing conditioning notation, is:

$$Pr(y_{ij^*} = 1) = \sum_{M \subseteq \mathcal{M}: j^* \in M} Pr(M_i = M) Pr\left(EU_{ij^*} > EU_{ik} \; \forall \; k \in M\right) \tag{3.2}$$

Each plan appears in an individual's consideration set with probability φ_j. Plan consideration probabilities are homogeneous across agents facing the same feasible choice set.[41] Conditional on observables, each plan's appearance in a consideration is assumed independent.[42] By independence, the probability of any consideration set $M_i = M \subseteq \mathcal{M}$ can be written in terms of the individual plan consideration probabilities:

$$Pr(M_i = M) = \prod_{k \in M} \varphi_k \prod_{k' \notin M} \left(1 - \varphi_{k'}\right).$$

In such a model, it is possible for an individual to draw an empty consideration set, $M_i = \varnothing$. In such cases, a simple completion rule is needed, such as those discussed in Barseghyan et al. (2019b). For simplicity this additional component of probability is left implicit, but in the event an individual draws $M_i = \varnothing$, I assume one of the 46 available plans is chosen with equal probability. The probability any beneficiary i selects plan j^* can be written as:

$$Pr(y_{ij^*} = 1) = \sum_{M \subseteq \mathcal{M}: j^* \in M} \prod_{k \in M} \varphi_k \prod_{k' \notin M} \left(1 - \varphi_{k'}\right) Pr\left(EU_{ij^*} > EU_{ik} \; \forall \; k \in M\right) \tag{3.3}$$

[41] This assumption can be relaxed by allowing φ_j to depend on beneficiary attributes.

[42] As noted below in Section 3.6.3 consideration probabilities are modeled as functions of plan attributes. Therefore specific forms of correlation between the consideration of similar plans is permitted.

As written, equation 3.3 requires enumeration of all possible consideration sets M. In a setting such as Medicare Part D where beneficiaries in California have 46 plans available, such an enumeration is computationally unfeasible. Rather than approximate such a sum with simulation of consideration sets, as done in Goeree (2008), this choice probability can be simplified to fully avoid the need to account for every potential consideration set. The utility model in equation 3.1 does not include an error term, and at any given value of risk aversion ν_i, all plans can be ranked by expected utility. That is, fix $\hat{\nu}$, and order plans from worst to best in terms of expected utility $EU_{i1} < EU_{i2} < ... < EU_{ij^*} < EU_{ij+1}... < EU_{iJ}$. Therefore, for plan j^* to have been selected at $\nu = \hat{\nu}$, the consideration set must not have included (at the minimum) plans $j+1, ..., J$, since if those plans were present, j^* would not be selected. Let $k \succ_{\hat{\nu}} j^*$ denote the set of plans that dominate j^* at a given value $\hat{\nu}$. Thus, conditional on $\nu_i = \hat{\nu}$, $Pr\left(EU_{ij^*} > EU_{ik} \; \forall \; k \in M\right) = 0$ if M contains any plans in the set $k \succ_{\hat{\nu}} j^*$ and $Pr\left(EU_{ij^*} > EU_{ik} \; \forall \; k \in M\right) = 1$ if M does not contain any plans $k \succ_{\hat{\nu}} j^*$.

Such a ranking and collection of dominating plans can be computed at any values of $\nu \in [0, .\bar{\nu}]$, where $\bar{\nu}$ is the upper bound on the coefficient of absolute risk aversion. Using this simplification, equation 3.3 for a given value of ν_i can be written without regard for specific consideration set as:

$$Pr(y_{ij^*} = 1 | \nu_i = \hat{\nu}) = \varphi_{j^*} \prod_{k \succ_{\hat{\nu}} j^*} (1 - \varphi_k) \tag{3.4}$$

These sets of dominating plans can be computed for each individual at any value of risk aversion. Using the Riemann approximation procedure described below, averaging equation 3.4 across individuals allows for approximation of

the choice probabilities of the form:

$$Pr(y_{j*} = 1) \int Pr(y_{j*}|\nu) dF_{\nu}. \tag{3.5}$$

3.6.3 Estimation

Consideration Probabilities

Beyond the utility-relevant variables governing the distribution of costs a beneficiary faces under each plan, I allow choices to depend additionally on non-monetary and monetary attributes through consideration. Consideration is modeled to depend on the insuring firm, the deductible, whether a plan offers any gap coverage, the count of 100 most popular drugs covered, and the average cost-share in the initial coverage phase.[43] To account for the higher weight placed on premiums relative to out-of-pocket costs in reduced form regressions, the plan's premium is also included as a determinant of consideration. Each plan's consideration probability, φ_j, is modeled as a function of these characteristics:

$$\varphi_j = f(\text{firm}_j, \text{premium}_j, \text{deductible}_j, \text{gap}_j, \text{Top100}_j, \text{AvgCS}_j) \tag{3.6}$$

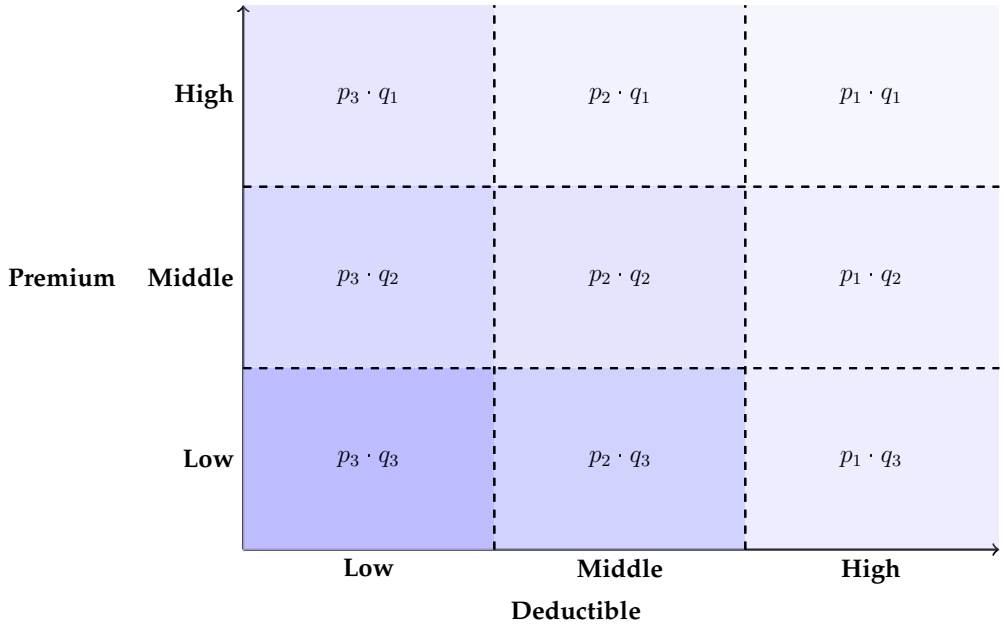
For intuition on how attributes relate to consideration, it is helpful to consider only two attributes: deductible and premium. An individual may limit

[43]The standard CMS plan includes a 25% cost-share in the initial coverage phase, but the realized cost-share can differ substantially due to modified plan structure, formulary, and differing drug needs across individuals. After estimating expenditures in every plan under perfect foresight, I compute the average ratio of out-of-pocket costs to gross expenditure within the initial coverage phase across all individual's whose out-of-pocket spending was between the deductible and initial coverage limit. Accordingly, this variable takes the same value for all beneficiaries.

their consideration of plans based on these attributes for a number of plausible reasons. The online tool CMS provides to aid in plan choice, PlanFinder, allows for plans to be sorted based on premiums or deductibles. This may lead an individual to only see the plans with the lowest premiums or deductible. In the presence of liquidity constraints, an individual may be unable to afford large lump expenses and only consider plans with reduced or eliminated deductibles. If an individual faces budget constraints, they will not consider plans with monthly premiums in excess of a reservation price. In a more general model of limited consideration, individuals are modeled as simply less likely to consider a plan with less desirable observable attributes than one with better attributes. There is an unambiguous ordering of both deductible and premium from most to least preferred. A lower deductible (premium) is clearly better than a higher deductible (premium), all else equal.

Figure 3.4 visually demonstrates the connection between the desirability of a plan attribute and consideration probabilities. In this example, deductible and premium each take one of three values: low, medium, or high. The probability of considering a low deductible plan (p_3) is higher than the probability of considering a medium deductible plan (p_2), which is higher than the probability of considering a high deductible plan (p_1). A similar ordering of consideration follows for premium. The bottom left region corresponds to the best plans along these two attributes - those with the most preferred low deductible and the most preferred low premium. The darker shade of blue reflects the largest consideration probability of these plans. As you move away from the bottom left of the figure, plans become increasingly less desirable along these attribute dimensions. Consideration is modeled to diminish as plans move further and further away from best along each attribute dimension. The lightest shaded box

in the upper right corner corresponds to plans with both the highest premium and highest deductible and are, thus, least likely to appear in an individual's consideration set.



Note: $p_1 < p_2 < p_3$ represents the consideration probabilities for each deductible and $q_1 < q_2 < q_3$ the consideration probabilities for premiums.

Figure 3.4: Consideration Intuition with 2 Attributes

With the exception of the insuring firm, all of the included attributes have such an objective ranking. All else equal, lower deductibles, premiums, and cost-sharing is preferred. Similarly covering more drugs is preferred to fewer, and gap coverage is better than no gap coverage. The intuition of Figure 3.4 is applied across these multiple dimensions. Although the illustrative example was a simplification, the idea of such diminishing consideration is appealing and converges to a specification of consideration that reflects a geometric decay of consideration probabilities as plans progressively become less and less desirable in their attributes. In the absence of an objective ranking over firms, and to reflect the numerous underlying mechanisms causing individuals considering

firms differentially, I model a base consideration probability for each of the 19 firms in the market. It is to this base probability that the reductions in consideration according to attributes is applied. The details of the parameterization is discussed in Appendix B.2.

Maximum Likelihood

In practice, the integral in equation 3.5 is estimated through a Riemann integral approximation. The support of the coefficient of risk aversion, $[0, \bar{\nu}]$, is divided into a fine grid. At each value of ν on the grid, for each individual, the set of plans $k \succ_\nu j^*$ is computed, as described in equation 3.4. To approximate the integral over the distribution of ν, I weight the choice probabilities above at each value of ν in the grid based on the probability density function of risk aversion at those grid values. Weighted individual choice probabilities are then logged and summed. I maximize the resulting loglikelihood to recover the values of the model parameters - including those governing the distribution of risk aversion - that best match the observed choices.

In all specifications, I assume the coefficient of absolute risk aversion is distributed according to a Beta distribution, $\nu \sim Beta(\beta_1, \beta_2)$. The Beta distribution is an appealing assumption due to its flexibility. Risk aversion is assumed to be bounded above by .01, a liberal assumption in light of Rabin (2000). Estimates are not sensitive to this assumption. See Appendix B.2 for more details on the procedure.

3.6.4 Identification

To separately identify consideration from risk preferences, I require a large support of certain variables and a form of an exclusion restriction. There must be sufficient variation in the utility-relevant variables to shift utility rankings of plans without correspondingly shifting consideration probabilities. In this model, the only utility-relevant variables are those governing the distribution of costs under each plan. Other plan attributes are presumed to impact consideration but not directly enter utility. The one variable that directly enters both utility and consideration is the premium of the plan. Barseghyan et al. (2019b) establishes that some overlap between variables in utility and consideration does not threaten identification provided there are other variables that shift either utility or consideration, but not both directly (i.e., an exclusion restriction). The consideration-relevant variables that relate to drug costs - the deductible, gap coverage, count of drugs covered, and realized average cost-share - impact the distribution of costs in a complex, highly nonlinear way. As a result, there is sufficient independent variation between the plan attributes and the utility-relevant variables to satisfy exclusion.

Identification can be viewed in two stages. First, to identify the consideration probabilities, φ_j, I require a large support for the utility-relevant variables, $\hat{\mu}_{ij}$ and σ^2_{ij}. Intuitively, there are regions of the support of these variables where certain plans are unambiguously best, regardless of risk preferences. Under full consideration, I would expect to see all individuals in that region of the support choosing the best plan. The empirical share of individuals selecting the plan in that region of the support identifies the consideration probability for that specific plan. Such an exercise can be repeated throughout the large support to

identify all of the φ_j probabilities. Variation of plan attributes within and across firms identifies the consideration effects of individual plan attributes. The second step is to identify the distribution of risk preferences. With consideration identified, this proceeds in the same manner as a full consideration model, as described in Matzkin (2007). Large variation in the mean and variance of costs traces out the distribution of ν among the population.

3.7 Results

3.7.1 Limited Consideration

The model of expected utility with limited consideration fits the data patterns of the California beneficiaries well. Heterogeneity in consideration sets plays a crucial role in prescription drug insurance choice. Table 3.4 presents the estimates of the impact of plan attributes on consideration. All parameters included are between 0 and 1 to bound corresponding consideration probabilities. The δ estimates reflect the total decay in consideration that occurs as the attribute progresses from the very best to the very worst. All else equal, the estimate for δ_{prem} indicates a plan with the highest premium is considered only 10% as much as the lowest premium plan. Similarly, a plan with the maximum deductible of $310 is considered 18% as much as an equivalent zero deductible plan. And plans lacking gap coverage are considered 86% as frequently as one with gap coverage.

The results in Table 3.4 are intuitive in a number of ways. As modeled, a plan's count of top 100 drugs covered and the average cost-share in the initial

Table 3.4: Model Results: Consideration Impact of Plan Attributes

	Estimate	95% CI
δ_{prem}	0.100	[0.074, 0.140]
δ_{ded}	0.182	[0.163, 0.206]
δ_{gap}	0.859	[0.782, 0.953]
δ_{top100}	1.000	[0.999, 1.000]
δ_{avgcs}	1.000	[1.000, 1.000]

Notes: All δ terms are defined between 0 and 1 and reflect how much consideration a plan with the worst value of an attribute is considered relative to an equivalent plan with the best value of the attribute. Confidence intervals based on 1,000 bootstraps with sub-sampling to correct for estimates on the boundary.

coverage phase do not impact its probability of consideration. These plan attributes are generally not immediately knowable to a beneficiary. An individual can find whether certain drugs are covered in a plan's formulary through tools such as Medicare's PlanFinder online tool, but a full count of such coverage of the 100 most popular drugs among beneficiaries is not published. Additionally, an astute beneficiary that seeks out extensive information on their plans can learn the copay and coinsurance rates for different tiers of drugs in the initial coverage phase - and in fact, that information is what the average cost-share variable is meant to proxy for - but such a precise aggregate measure is not feasible to compute for most individuals. Moreover it is computed based on all of the beneficiaries' drug needs, which is also not information any individual beneficiary has at the time of plan comparison. To the extent this captures

behavioral trimming of choice sets based on desirable attributes, it is not particularly surprising that these more difficult to ascertain attributes are not strong drivers of consideration. The first three attributes, in contrast, are unambiguous and more easily known to beneficiaries. In fact, many online tools, including the PlanFinder, summarize this information for beneficiaries. As mentioned, individuals can even sort and filter available plans on the PlanFinder by premiums and deductibles.[44]

In addition to the impact of these attributes, the probability any plan is considered is largely determined by which firm offers the plan. Figure 3.5 presents the insuring firm base consideration probabilities. Three large firms, UnitedHealth, Blue Cross of California, and Anthem, garner near full base consideration. Each of these firms offers a plan included in the 5 most chosen plans within the sample. Nearly half of the 19 firms in the market are considered with probability below 10%, even before accounting for the impact of plan attributes. Such heterogeneous consideration across firms may reflect, among other explanations, the impact of differential advertising, agent steering effects, or enrollees' insurance experiences prior to Medicare. I leave to future research the detailed analysis of such explanations.

My estimates capture substantial heterogeneity in consideration sets across beneficiaries. Figure 3.6 presents the implied distribution of consideration set sizes across individuals in the sample. Although the market includes 46 plans, consideration sets do not come even close to including that many plans. The vast majority of beneficiaries consider no more than 5 plans, and no one is estimated to consider a set containing more than 14 plans. Approximately 14% of

[44]Following the Affordable Car Act, Medicare plans no longer include the coverage gap. As such, today's PlanFinder does not present this information.

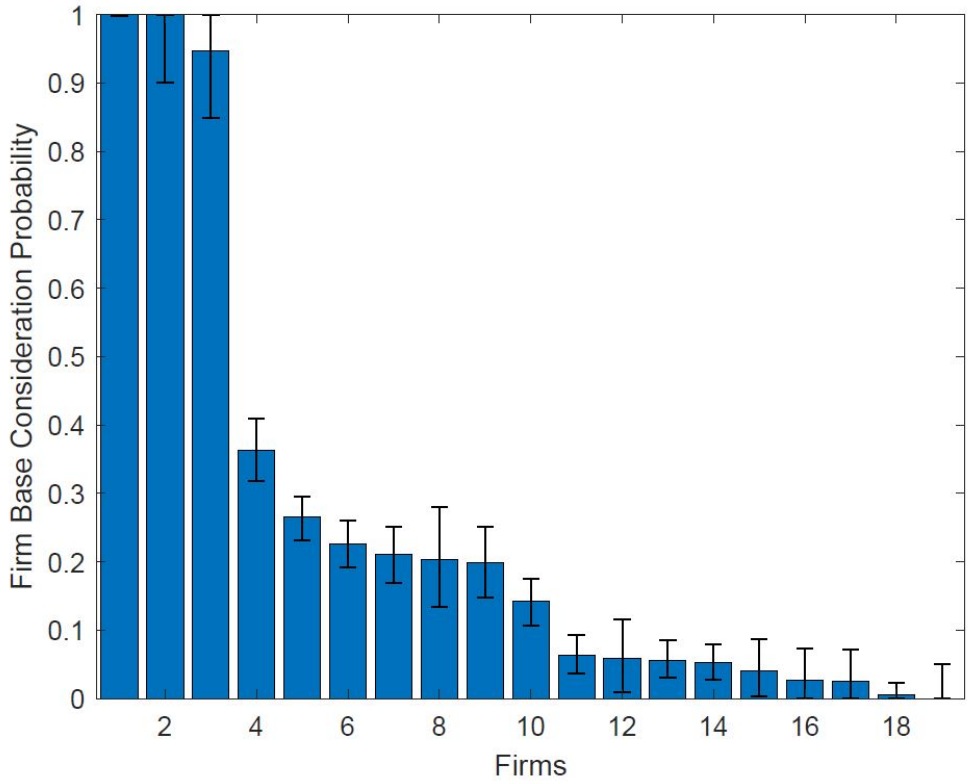


Figure 3.5: Model Results: Firm Base Consideration Probabilities

Notes: Firms are ordered based on estimated base consideration probabilities. Error bars present 95% confidence intervals based on 1,000 bootstrap repetitions with sub-sampling to adjust for estimates on the boundary.

individuals evaluate a single plan. The composition of these consideration sets is highly concentrated among the plans that share the most popular attributes. As shown in Figure 3.7, the largest firms account for an overwhelming share of the plans considered. The three large firms described above constitute over 60% of all plans considered. The three firms with the smallest firm effects account for fewer than 0.5% of plans considered. In fact, 7 of the 19 firms each represent fewer than 1% of considered plans and cumulatively represent just over 2% of all plans considered. These plans, although infrequently considered and chosen, are nonetheless good plans for a nontrivial portion of the sample.

As shown in the contrast of the blue shares of considered plans and the red shares of the feasible menu, this pattern is not an artifact of the number of plans offered, but rather, reflects the strong positioning of a few large firms.

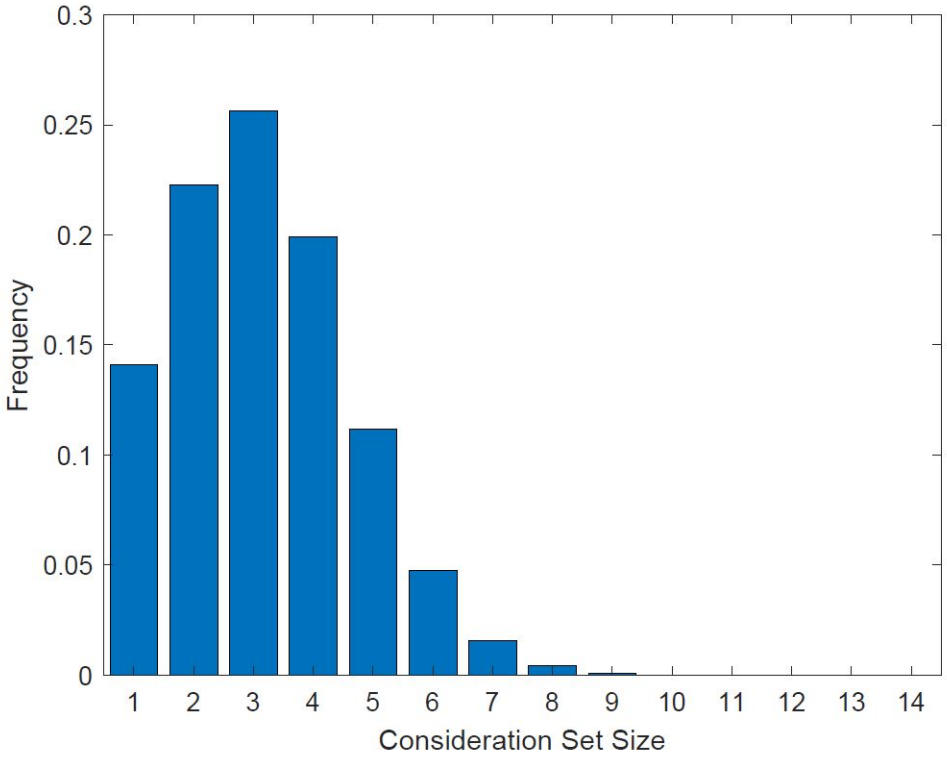


Figure 3.6: Implied Distribution of Consideration Set Size

Notes: Consideration set sizes estimated as the average over 1,000 simulations of individual risk aversion and consideration sets for the analysis sample.

Consideration sets are similarly skewed towards zero deductible plans. Plans without a deductible account for 19 of the 46 plans in California in 2010 but nearly 75% of considered plans.[45] Figure 3.8 gives a simple illustration of this pattern in the first panel. The pattern of the premiums of considered plans is more nuanced. The second panel of Figure 3.8 plots the share of considered plans based on bins of premiums. The first bar represents the 10 lowest premi-

[45]Moreover, 95% of beneficiaries consider at least one plan with a fully eliminated deductible.

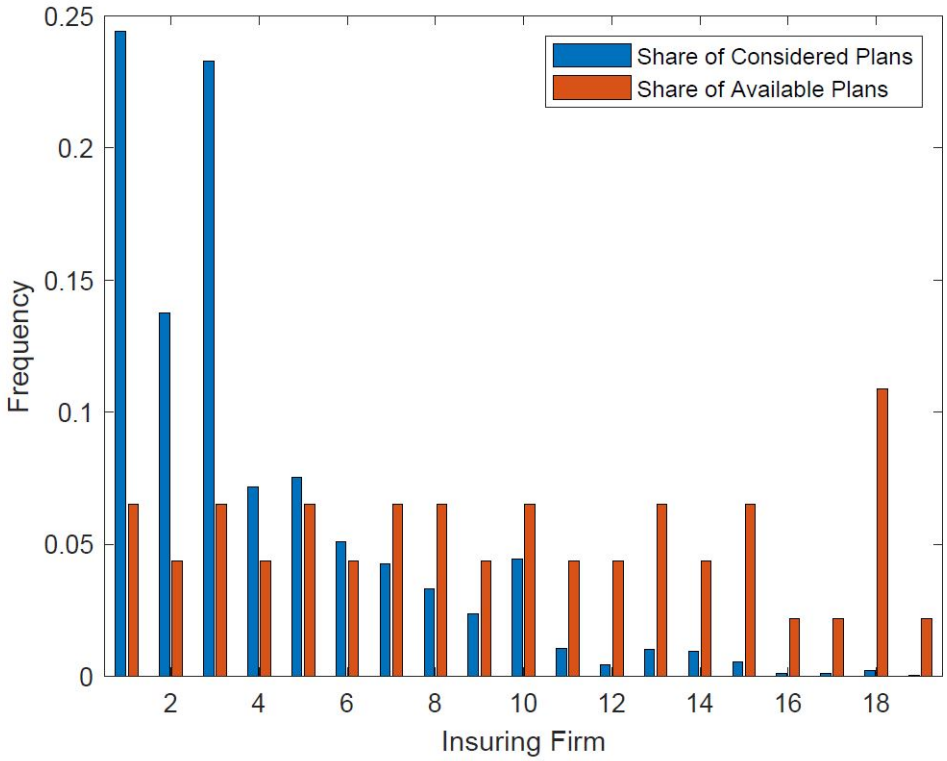
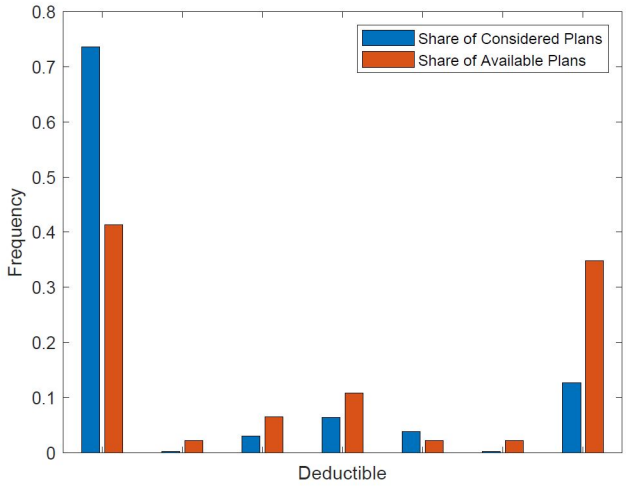


Figure 3.7: Implied Shares of Consideration Sets and Choice Set by Firm

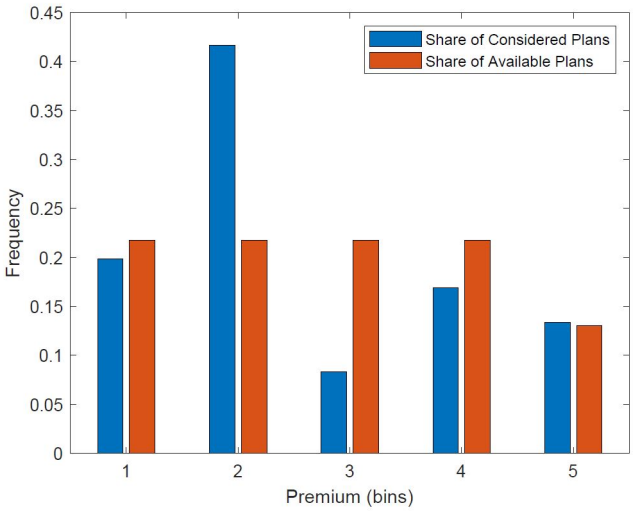
Notes: Firms are ordered as in Figure 3.5 based on estimated firm base consideration probabilities. Shares of consideration sets are based on 1,000 simulations of individual consideration sets for the analysis sample.

ums, the second bar the next 10 lowest premiums, and so on. While the estimate of δ_{prem} conforms with the intuition that higher premium plans are considered less than more appealing lower premium plans, the plans with the lowest premiums are generally those with higher deductibles. Thus, this preference towards lower premium plans alongside low deductibles manifests in the plans in the second bin of premiums accounting for a disproportionate share of plans considered.

The resulting consideration probabilities of the 46 plans vary substantially.



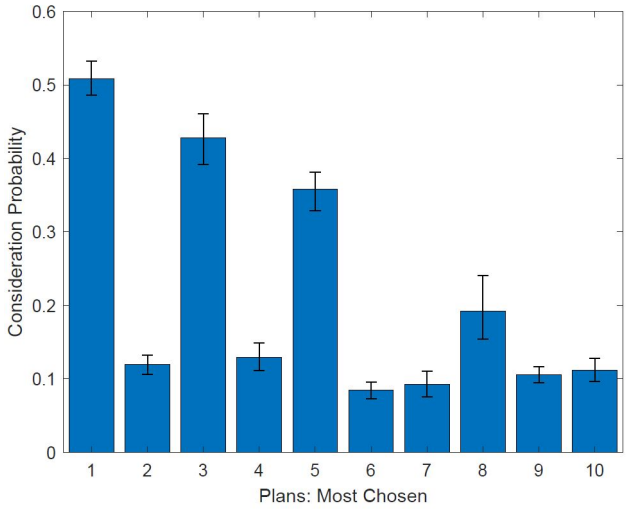
(a) Share of Consideration Sets by Deductible



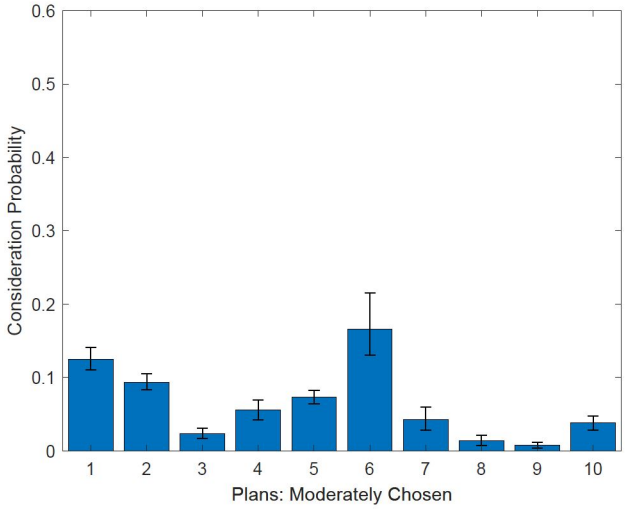
(b) Share of Consideration Sets by Premium (bin)

Figure 3.8: Implied Shares of Consideration Sets and Choice Set by Deductible and Premium

Notes: Panel (a) is ordered left to right from $0 to $310 deductibles. Premiums in Panel (b) are ordered lowest to highest by bins of 10. Shares of consideration sets are based on 1,000 simulations of individual consideration sets for the analysis sample.



(a) 1-10 Most Popular Plans



(b) 11-20 Most Popular Plans

Figure 3.9: Plan Consideration Probabilities of 20 Most Chosen Plans

Notes: Panel (a) presents the model implied consideration probability for the 10 plans with the largest shares of enrollment, with the most chosen plan first and the 10^{th} most chosen plan last. Panel (b) presents the same information for the 11^{th} through 20^{th} most chosen plans. Error bars present 95% confidence intervals based on 1,000 bootstrap repetitions with sub-sampling to adjust for estimates on the boundary.

Figure 3.9 presents consideration probabilities, φ_j as described in Equation 3.6, for the 20 most popular plans. Plans are ordered based on empirical choice shares. Even among these relatively popular plans, consideration probabilities are frequently modest. Figure 3.12 below shows how these consideration probabilities bridge the gap between observed plan choices and those implied by risk preferences under full consideration. The results on consideration are consistent with a number of underlying sources of limited consideration. The strong impact of the deductible on consideration coheres to stories of liquidity constraints, a reality for many Americans, as noted in Durante and Chen (2019). It is both plausible and rational for such a constrained beneficiary to consider exclusively, or nearly exclusively, plans with an eliminated deductible, as my estimates indicate. The result that the count of top 100 drugs covered does not impact consideration presents a lack of evidence of filtering on drug plans based on formulary generosity, or at least using such a general measure of formulary generosity. The substantial role of firm effects in consideration lends support to a number of behavioral forces resulting in limited consideration. Familiarity of firms based on prior insurance experience or social influence, arising the insurer of friends or spouses, can lead beneficiaries to filter according to preferred firms. There is substantial firm advertising in this market and these results may reflect the consideration impact of advertising campaigns.

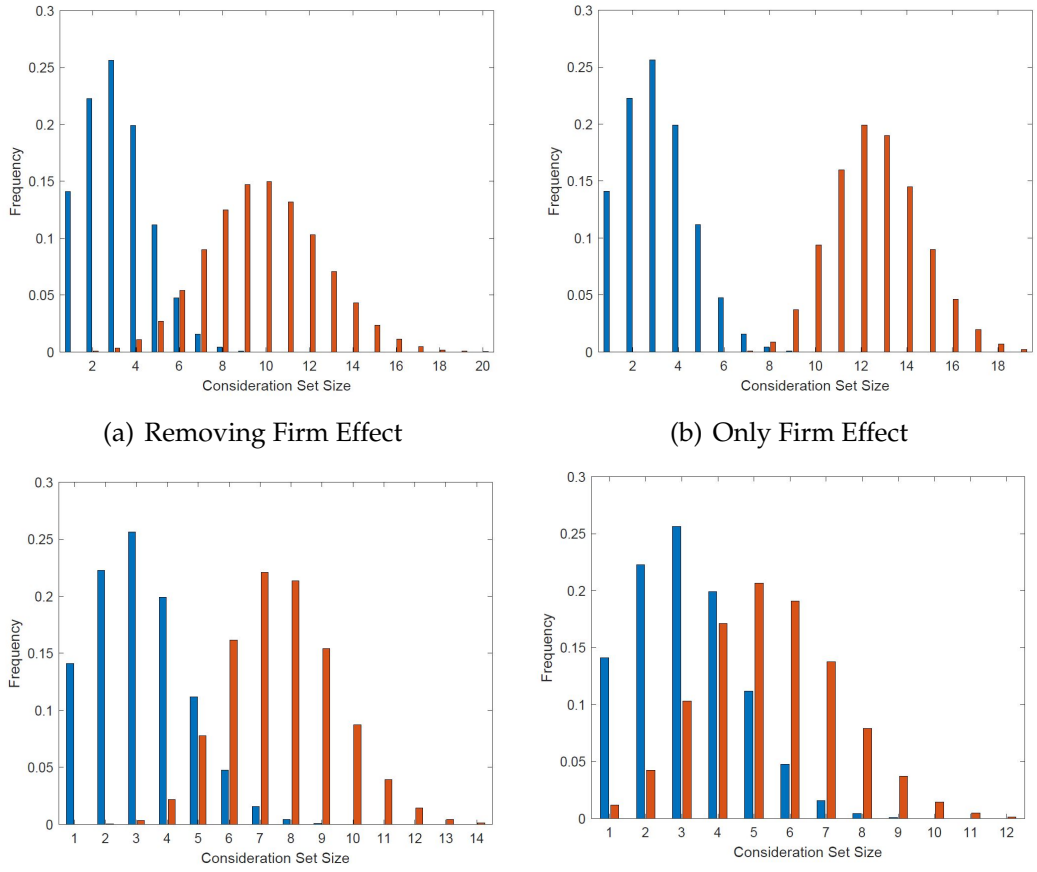
The cumulative impact of these attributes on consideration results in consideration sets that are much smaller in size than the feasible choice set. The modal consideration set contains 3 plans and over 90% of beneficiaries consider a set with 5 or fewer plans. To parse the effects of each attribute on the resulting consideration set composition, I simulate consideration sets when certain impacts of consideration are eliminated. Figure 3.10 plots the distribution of

consideration set sizes across these counterfactual schemes. Holding all other estimates fixed, Panel (a) presents the impact on consideration set size when the firm effect is eliminated. In practice, this exercise translates to assigning all firms a base consideration probability of 1. This alteration results in a rightward shift of the distribution of consideration set size, as fewer plans are immediately eliminated as a result of firm filtering. Panel (b) presents the opposite exercise where the firm effect is the sole determinant of consideration. Because the three largest firms have base consideration probabilities of 1, or nearly 1, by construction consideration sets have a larger minimum number of plans. The bottom two panels reflect the elimination of the deductible and the premium from consideration, generating a more modest increase in the size of consideration sets.

3.7.2 Risk Preferences

In contrast to the previous literature on plan choice in Medicare Part D, I find estimates of risk aversion among California's beneficiaries comparable to other insurance settings. Table 3.5 describes estimates of risk preferences in the sample. The top panel provides the mean and variance of risk aversion in the model with limited consideration. The estimate of mean risk aversion is on par with previous studies that use field data to measure risk preferences, and comes along with moderate variance.[46] These estimates can be difficult to interpret and compare without additional context. Table 3.5 includes a measure of risk premium for an individual with CARA utility facing a lottery that results in a

[46]For example, see Barseghyan et al. (2019a),Handel and Kolstad (2015), Handel (2013), Barseghyan et al. (2013), and Cohen and Einav (2007). In particular, Barseghyan et al. (2019a) finds that the incorporation of limited choice sets can rationalize auto collision choices with lower and more homogeneous of risk aversion than standard full consideration models. My estimates of the mean and variance risk aversion are on the lower end of the fairly narrow confidence intervals of that model.



(a) Removing Firm Effect

(b) Only Firm Effect

(c) Removing Deductible Effect

(d) Removing Premium Effect

Figure 3.10: Baseline and Counterfactual Distributions of Consideration Set Size

Notes: All subfigures present implied consideration sets sizes of the baseline estimates in blue and when the following adjustments to estimates are simulated in red: Panel (a) all firm base probabilities are set to 1; Panel (b) δ_{prem}, δ_{ded}, and δ_{gap} in Table 3.4 are all set to 1; Panel (c) δ_{prem} is set to 1; Panel (d) δ_{ded} is set to 1. Shares of consideration sets are based on 1,000 simulations of individual consideration sets for the analysis sample.

loss of $1,000 with 25% probability. Such a lottery has an expected value of a $250 loss. An individual with a coefficient of risk aversion equal to my baseline mean estimate would be willing to pay a risk premium of $102 to avoid such a lottery. In contrast, a standard CARA random expected utility model with full consideration substantially underestimates risk aversion, as shown in the lower panel of Table 3.5. A full consideration model with a constant coefficient of risk aversion across agents finds risk neutrality and matches choice patterns poorly. Incorporating heterogeneity in risk aversion increases the suggested levels of risk aversion, but even the upper bound of the confidence interval in such a model falls below the lower bound of the confidence interval on mean risk aversion under limited consideration.

Table 3.5: Model Estimates: Risk Preferences

	Estimate	Risk Premium	95% CI
Limited Consideration			
$\mathbb{E}$(Risk Aversion)	$9.52 \cdot 10^{-4}$	$102	$[5.59 \cdot 10^{-4}, 1.40 \cdot 10^{-3}]$
Var(Risk Aversion)	$3.14 \cdot 10^{-6}$		$[9.75 \cdot 10^{-7}, 6.31 \cdot 10^{-6}]$
Comparison - CARA RUM Full Consideration			
Homogeneous Risk Aversion	$1.33 \cdot 10^{-7}$	$0	$[7.28 \cdot 10^{-8}, 1.33 \cdot 10^{-7}]$
Heterogeneous $\mathbb{E}$(Risk Aversion)	$4.28 \cdot 10^{-4}$	$43	$[3.78 \cdot 10^{-5}, 4.60 \cdot 10^{-4}]$
Var(Risk Aversion)	$1.71 \cdot 10^{-6}$		$[1.14 \cdot 10^{-8}, 2.17 \cdot 10^{-6}]$

Notes: CI based on 1,000 bootstraps. In limited consideration model, sub-sampling used to correct for estimates on the boundary. Risk premium is calculated for a beneficiary facing a lottery that results in a loss of $1,000 with 25% probability.

The inclusion of a role for these non-monetary and non-cost attributes is important for estimating risk aversion in this setting. The assumption of full consideration results in an underestimation of risk aversion. The financial stakes in the market for prescription drug insurance are relatively modest for most beneficiaries, although certainly not trivial. In other insurance settings, however, the

financial implications are enormous. Failing to account for consideration and the resulting estimates of a misspecified choice model limits the usefulness of the model for understanding consumer behavior or policy implications. While the way in which biased estimates of risk aversion would naturally imply misspecified demand for insurance is well understood, the material importance of limited consideration is not.

3.7.3 Plan Choice under Limited Consideration

Choices in this market are driven both by risk aversion and limited consideration. Taken together, the model of expected utility with limited consideration matches the observed choice patterns of beneficiaries well. Figure 3.11 plots the implied choice distribution of the baseline model alongside the empirical distribution of plan choices. The right-most bar of the figure is a composite plan comprised of all the plans for which between 1 and 10 beneficiaries in the analysis sample enrolled.[47]

By estimating risk preferences without the distorting impact of other plan attributes, I am able to compute the model implied choice distribution under full consideration. For the 20 plans with the largest empirical enrollment shares, Figure 3.12 plots the implied distribution under limited consideration in blue, which as seen in Figure 3.11 is similar to the empirical distribution. The implied full consideration choice shares are shown next in red. On the whole, these two distributions differ quite substantially. The bridge between the two distributions is consideration. The third bar plotted in yellow is the model im-

[47] As a privacy measure, CMS has a cell suppression policy that prevents the release of statistics such as choice frequencies if fewer than 11 individuals underlie the statistic.

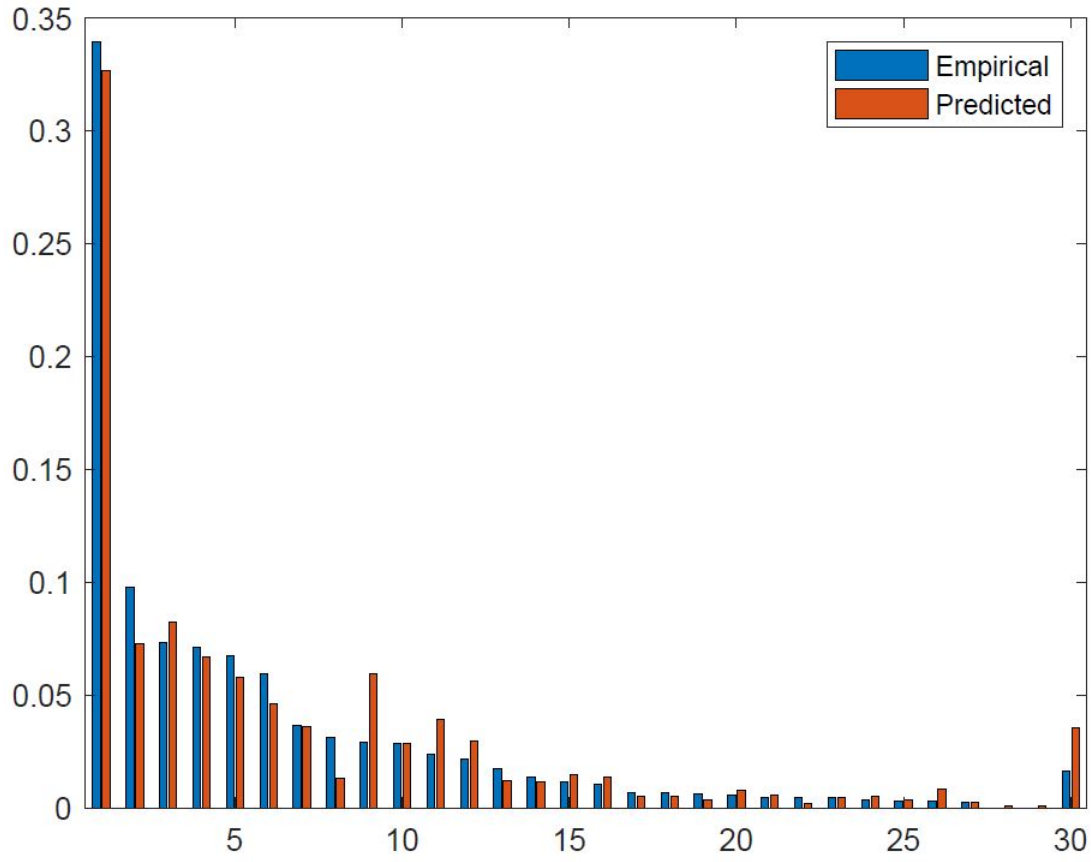
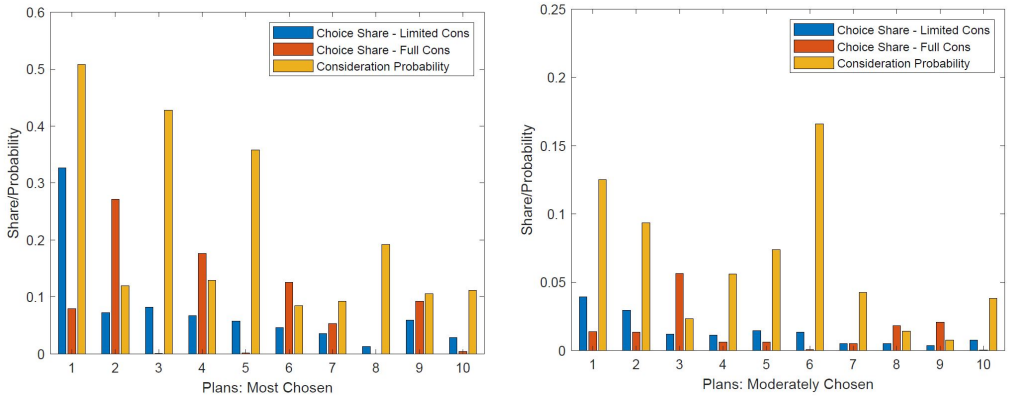


Figure 3.11: Empirical and Model Choice Distributions

Notes: The blue bars correspond to the empirical choice shares, and the red bars are the implied choice distribution based on 1,000 simulations. Plans are ordered from the plan with the largest enrollment share on the left to the plans with zero enrollment. The rightmost plan corresponds to a composite plan of the 17 plans in which between 1 and 10 individuals enrolled.

plied consideration probability for each plan. Plan consideration probabilities are large for those plans that are chosen relatively frequently but are optimal rather infrequently under full consideration. Similarly, plans that are chosen infrequently in practice compared to under full consideration correspond to relatively smaller consideration probabilities.



(a) 1-10 Most Chosen Plans (b) 11-20 Most Chosen Plans

Figure 3.12: Choice and Consideration Probabilities: Limited and Full Consideration, 20 Most Chosen Plans

Notes: Plans are ordered left to right by empirical enrollment shares. Implied choice shares are estimated as the average over 1,000 simulations of individual risk aversion and consideration sets for the analysis sample.

As a result of limited consideration, individuals cluster on plans with low deductibles offered by a few popular firms. Many of these plans, however, are not particularly well suited to the drug needs and risk preferences of many. The discrepancy between the two choice distributions in Figure 3.12 is caused by many beneficiaries considering a set of plans that does not contain their best plan. As a measure of the cost of limited consideration in this population, I compute the difference in certainty equivalent of the chosen plan and the optimal plan. The average certainty equivalent loss across the sample in my model is $226. Table 3.6 compares the average certainty equivalent difference under the same counterfactual exercise in Figure 3.10 in which consideration effects are eliminated in simulations, maintaining all other estimates. The reduction in the average difference in certainty equivalent between chosen and optimal plans arises from the corresponding increase in consideration set sizes displayed in Figure 3.10, and the increase in individual plan consideration probabilities. The

changes in certainty equivalent difference highlight the sizable role of attributes in consideration and the cost of limited consideration.

The values in Table 3.6 admit the following two interpretations. Since the model of limited consideration nests many behavioral models of consideration set formation, the model is agnostic about why the attributes determine consideration sets. Take, for example, the firm effect. First, the impact of firm identity on choices can be purely a consideration impact; advertising by some firms in this market is substantial. In that case, the difference between the baseline certainty equivalent loss of $226 and the $88 loss under a counterfactual without the firm effect represents meaningful welfare improvement. By reducing consideration based on firm identity, beneficiaries are on average losing $138. Second, if the impact of firm on plan consideration represents unobserved quality, then a counterfactual that removes an individual's ability to filter plans considered based on the firm would be welfare reducing. In that case the $138 additional lost certainty equivalent can be interpreted as an average bound on the shadow price of unobserved quality.

A similar logic applies to the other counterfactual results. If the role of the deductible is through consideration, then removing this channel and the corresponding increase in consideration set sizes improves welfare.[48] Individuals save on average $85 in that counterfactual. However, if the role of the deductible in consideration represents liquidity constraints, then individuals will become worse off without the ability to limit plans based on deductible. For the deductible, premium, and gap effects, if the impact my model captures is consideration rather than constraints, the changes in average certainty equivalent can

[48]For example, available plans in the CMS sponsored online PlanFinder tool can be sorted by deductible. It is plausible an individual presented with plans in that order does not look beyond the initially presented plans and only considers zero or low deductible plans.

be interpreted as clear welfare improvements through reducing consideration obstacles. If, however, these effects are manifestations of binding constraints, these changes represent the shadow prices of said constraints.

Table 3.6: Baseline and Counterfactual Results: Average Certainty Equivalent Loss Due to Limited Consideration

	Average CE Difference
Baseline Model	$226
Removing Firm Effect	$88
Only Firm Effect	$84
Removing Deductible Effect	$141
Removing Premium Effect	$147
Removing Gap Effect	$201

Notes: Average CE Difference computed as the average difference in certainty equivalent of the chosen plan and the optimal plan over 1,000 simulations of risk aversion and consideration sets. The counterfactual values are computed similarly when consideration sets are simulated with different components of plan consideration probabilities set to 1.

There has been discussion since Part D's inception that the large number of available plans is unwieldy even for the most sophisticated of enrollees. Changes to market regulations over time have reduced the number of available plans from approximately 50 in the early years of the program to closer to 30 plans presently. Without estimates of limited consideration, not to mention risk preferences, it is difficult to assess how a reduction in plans available will affect beneficiary choice quality. To illustrate this point, I conduct two counterfactual simulations that resemble CMS policies aimed at reducing the number of plans in the market. A priori, the welfare impact of such a policy is ambiguous. The

removal of plans may harm individuals who are no longer able to select their preferred plans. However, the reduction in the size of the feasible choice set may remove distracting plans. In the first exercise, I simulate choices when the set of feasible plans excludes plans with few enrollees. This exercise is inspired by the policy position of CMS beginning in 2010 that recommends plans with few enrollees consolidate or exit the market. In my analysis sample, the elimination of all plans that do not meet an enrollment threshold of 0.5% of the sample corresponds to a reduction in the choice set from 46 to 20 plans.[49] The consideration sets continue to focus on a small set of plans and largely do not change. The 26 eliminated plans are both infrequently considered and infrequently optimal, and thus do not alter the certainty equivalent comparison substantially. Within the market of 20 plans, the average certainty equivalent difference of the chosen and optimal plans remains $226. When comparing the simulated chosen plans to the optimal under the original menu of 46 plans, the average certainty equivalent differences increases slightly to $230 due to the small number of beneficiaries whose optimal plans are no longer available.

The second counterfactual limits the number of plans an individual firm can offer. In 2010, CMS also introduced a meaningful differences requirement. In subsequent years, firms that offered multiple enhanced plans were required to establish a measurable difference in the expected costs of those plans.[50] As a result, most firms began to offer only two plans within a market.[51] I conduct a counterfactual where each of the 19 firms offers up to 2 plans.[52] This results

[49] An enrollment threshold of 1% would lead to a choice set of 16 plans.

[50] Enhanced plans refer to plans that offer enhanced features relative to the standard base plan.

[51] CMS has recently relaxed this regulation.

[52] Similar to the policy, I allow the firm to offer one standard deductible plan without gap coverage and one enhanced plans. In the case a firm offers multiple plans of one type, I retain the plan with a higher empirical choice share.

in a feasible menu of 35 plans. Similarly to the previous exercise, this leaves the average certainty equivalent largely unchanged. Although there is effectively no impact on consumer outcomes without these additional plans, insofar as providing these plans is not costless to firms, removing them from the market makes economic sense. However, with the understanding that beneficiaries exhibit substantial limited consideration, many efforts to streamline the market do not improve consumer outcomes. These results suggest, rather, that policymakers seeking to push beneficiaries towards better plans may want to instead encourage firms to compete on or implement regulation to standardize plans along consideration relevant attributes. Without an understanding of true risk preferences and how certain plan attributes impact consideration, it is difficult for policymakers to determine the impacts of market regulations or design changes.

3.7.4 Comparison to Standard Models

The workhorse model of insurance demand is the expected utility model with full consideration. Individuals are assumed to derive utility based on their risk aversion and the distribution of monetary outcomes under each available plan. Similar to the model under limited consideration presented in this paper, the only utility-relevant variables are those governing the distribution of losses. Non-monetary attributes are not provided a role in the decision framework. Beginning with the expected utility specification in Equation 3.1, I estimate a random utility model for comparison and to highlight the empirical advantages of accounting for limited consideration.[53] In both the homogeneous and hetero-

[53]It is worth noting that Apesteguia and Ballester (2018) describe a theoretical shortcoming of the random expected utility model in the insurance setting due to the implied non-monotonicity

geneous risk aversion specifications, the utility error is assumed to be iid Type 1 Extreme Value distributed.

$$EU_{ij} = -\exp(\nu\hat{\mu}_{ij} + \frac{1}{2}\nu^2\sigma_{ij}^2) + \epsilon_{ij} \tag{3.7}$$

$$EU_{ij} = -\exp(\nu_i\hat{\mu}_{ij} + \frac{1}{2}\nu_i^2\sigma_{ij}^2) + \epsilon_{ij},\ \nu_i \sim Beta(\beta_1, \beta_2) \tag{3.8}$$

The resulting estimates of risk preferences are described in Table 3.5 above. The assumption of full consideration in the CARA expected utility model results in substantial underestimation of risk aversion. In a model of homogeneous preferences, beneficiaries are estimated to be effectively risk neutral, with an estimated risk premium for a 25% loss of $1,000 of a mere penny. This is a puzzling result in an insurance market, and the implied choice probabilities of this model come close to rolling a 46-sided die. The inclusion of random preferences also underestimates risk aversion relative to the model of limited consideration.[54] The omission of non-monetary, and non-cost more generally, attributes from a model of plan choice also diminishes the ability of the model to rationalize observed choice patterns. The empirical fit of the heterogeneous model is shown in Figure 3.13. The most popular plan is markedly under-predicted and the small plans within the composite bar on the far right are largely over-predicted.

Acknowledging the importance of non-monetary attributes in rationalizing the choice of prescription drug plans, previously used methods take the approach of adding the plan attributes directly into utility. This can be done by scaling up the attributes by a coefficient to estimate. This suggests an interpretation of the coefficient as translating the variable into a "cost", comparable to

of choice probabilities in risk aversion.

[54] Although the estimate of mean risk aversion in the full consideration model here underestimate risk aversion, these estimates are still above those found in previous literature across all regions. It is possible that regional variation in consideration distorts further the estimation of risk preferences in the aggregate.

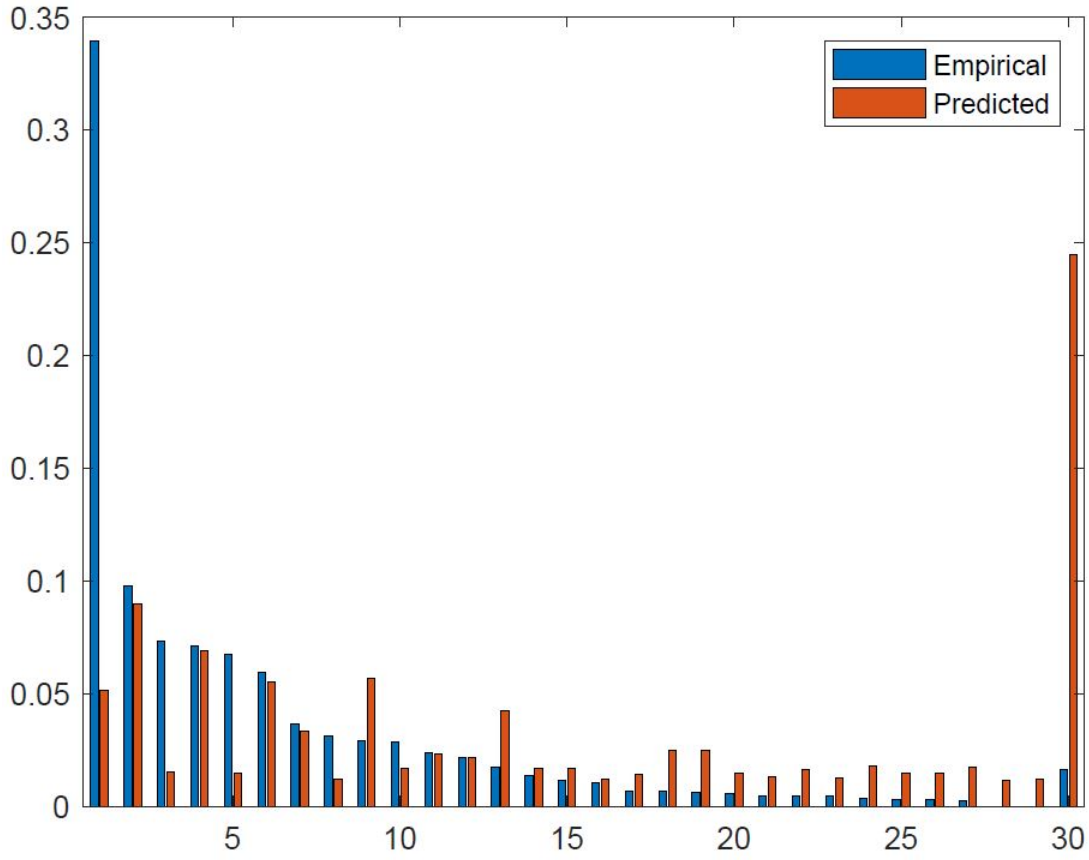


Figure 3.13: Implied Choice Distribution CARA Random Utility Model (Heterogeneous)

Notes: Plans are ordered based on empirical choice shares. Predicted choice probabilities are computed based on averaging of 1,000 simulations across the analysis sample.

the monetary attributes such as premium and out-of-pocket costs. Denoting the included non-monetary attribute by X_j and the monetizing scaling coefficient as γ, this translates in its simplest form here to the modified expected utility specification:

$$EU_{ij} = -\exp(\nu_i(\hat{\mu}_{ij} + \gamma X_j) + \frac{1}{2}\nu_i^2\sigma_{ij}^2) \tag{3.9}$$

In practice, expected utility is often estimated as a conditional logit by including the non-monetary attributes additively and a Type 1 Extreme Value

error.[55] In some applications, this can be a reasonable modeling assumption, but in the choice over prescription drug plans, this approach has certain undesirable features. One issue is the manner in which the linear approximation of utility typically abstracts away from more precise distributional estimates of risk preferences. More challenging, however, is the interpretation of estimates in this modified model. The estimates of coefficients on non-monetary attributes are generally interpreted relative to the coefficient on either premium or out-of-pocket costs as a willingness to pay for the attribute. Such a comparative interpretation is common in discrete choice models. This comes from modeling the incursion of the attribute as a utility "cost", on par with monetary costs. As discussed in Handel and Kolstad (2015) and Handel (2013), this approach treats the utility cost of the attribute as constant across the distribution of losses. Effectively, the utility cost of the attribute is a mean shift of the distribution of drug costs arising from uncertain drug needs. It can be difficult to attribute an economic meaning to these estimates. If the inclusion of a non-monetary attribute into utility is meant to capture a measure of non-financial plan quality or the impact of constraints such as liquidity constraints, it is not clear why that utility cost would be equivalent in the state of the world where an individual is healthy and does not file any drug claims and the state of the world where she is very ill and files many drug claims. Depending on the context this may or may not be of particular concern, but in this setting, it makes structural interpretation challenging.

This also raises questions regarding how to incorporate those estimates in a counterfactual analysis. The implications of such estimates in my analysis sample, specifically, do not conform with economic rationality. Table 3.7 presents

[55] See Abaluck and Gruber (2011) for a derivation of the conditional logit as a linear approximation of a CARA expected utility model.

estimates of the conditional logit with and without additional plan attributes. According to the estimates in Column (3), a dollar of deductible is equivalent to approximately $1.25 in premium and $1.92 in expected out-of-pocket costs. Taking these ratios, this would suggest that to reduce the deductible from the maximum allowed of $310 to $0, a beneficiary is willing to pay approximately $390 in premiums or $597 in expected out-of-pocket costs. Such estimates of WTP are wholly implausible and do not suggest an economic rationale for the estimated importance of the deductible in explaining plan choices. The monetary impact of the deductible is already accounted for in the expected out-of-pocket cost. As such, the result that the coefficient on deductible is statistically larger in magnitude than either premium or expected out-of-pocket costs in both Columns (3) and (4) is not consistent with the structural foundation of the model.

These counterintuitive results may be omitted from welfare analyses by assuming that the decision utility individuals use to select plans differs from their experienced utility. Under such an assumption, a researcher may be seen as estimating the foundational utility preferences driving behavior while accounting for the impact of such non-monetary attributes in choice. The inclusion of those variables in the estimation, however, can affect the primary estimates of interest. Columns (1) and (2) correspond to a standard insurance framework and include only utility relevant monetary variables. The inclusion of firm dummies and other attributes results in estimates of risk aversion approximately a third of the estimates without the attributes. The exercise to recover underlying preferences is sensitive to the inclusion of such additional variables.

Table 3.7: Plan Choice Estimates: Conditional Logit

	(1)	(2)	(3)	(4)
Premium	-0.5580	–	-0.536	–
	(.013)		(.026)	
$\mathbb{E}$OOP	-0.451	–	-0.350	–
	(.011)		(.012)	
Prem + $\mathbb{E}$OOP	–	-0.507	–	-0.381
		(.009)		(.011)
Variance	-0.026	-0.022	-0.007	-0.007
	(.002)	(.002)	(.002)	(.002)
Deductible	–	–	-0.674	-0.634
			(.026)	(.026)
Gap	–	–	-0.391	-0.767
			(.094)	(.076)
Top100 Drugs	–	–	-0.060	-0.071
			(.007)	(.006)
Avg CS	–	–	-0.757	-0.231
			(0.611)	(.611)
Firm Dummies	No	No	Yes	Yes
Implied Risk Aversion	$1.13 \cdot 10^{-3}$	$8.74 \cdot 10^{-4}$	$4.10 \cdot 10^{-4}$	$3.49 \cdot 10^{-4}$

Notes: Standard errors are in parentheses. Dollar denominated variables are measured in hundreds of dollars. Each column provides coefficient estimates from separate conditional logit maximum likelihood estimations. Variance denotes the variance of $\mathbb{E}$OOP measured in hundreds of dollars. Columns (3) and (4) include firm fixed effects. Corresponding risk aversion is computed by adjusting the coefficients on Variance and $\mathbb{E}$OOP (or Premium + $\mathbb{E}$OOP) for nominal dollars and taking the ratio of twice the adjusted Variance coefficient divided by the adjusted $\mathbb{E}$OOP coefficient.

3.8 Conclusion

This paper evaluates insurance choices in a setting where non-monetary plan attributes are suspected to influence plan choice using a theoretically appealing limited consideration framework that maintains the structure of expected utility. The motivation for this model comes from the inconsistency between standard models of insurance choice which describe insurance contracts by their monetary outcomes and the empirical correlation between insurance plan choices and other plan attributes. Since prescription drug coverage was introduced to Medicare in 2006, researchers have encountered challenges in rationalizing a sizable fraction of observed plan choices, as is the case in many other health insurance markets. Expected utility alone, the classic workhorse model of insurance choices, does not match the choices of beneficiaries well. Alternative methods of adding non-cost plan attributes, which are important for matching empirical patterns, into a utility framework result in estimates that are difficult to structurally interpret in this environment.

The model incorporates both non-monetary attributes and the effect of monetary attributes above and beyond their direct impact on beneficiary costs into the decision framework through limited consideration. Relaxing the standard assumption of full consideration, I model beneficiaries as expected utility maximizers over an unobserved subset of available plans contained in their consideration set. There are numerous plausible explanations for limited consideration in this market. The analysis sample of beneficiaries in California in 2010 faced a choice set of 46 different prescription drug insurance plans. Due to the advanced age and health conditions of the typical beneficiary, it is likely some individuals are unable to evaluate the entire choice set. Even for individuals lack-

ing cognitive limitations, the time required to consider and compare 46 plans may be too costly. These individuals may use certain plan attributes to trim the choice set down to a manageable size. The reduction of the choice set according to attributes may reflect unobserved constraints on an individual or the impact of firm advertising or the presentation of plans to the beneficiary. My model is agnostic about the underlying source of limited consideration but provides important insight into what features of plans drive consideration. Moreover, my model of consideration and the relationship of plan attributes to consideration, is computational tractable and is not subject to a curse of dimensionality as feasible choice sets increase. Since plan consideration probabilities are modeled as functions of plan attributes, it is the number of plan attributes, not the number of plans, that determines the number of parameters for estimation. This feature is especially appealing as the vector of estimated parameters may converge to a fixed number as the size of the choice set is increased. In a market with many choice sets, including the national PDP market, this is a very useful feature of the model.

My results show that heterogeneity in consideration sets plays an important role in plan choice. Despite the set of available plans, beneficiaries are largely considering no more than 5 plans. Which of the 46 plans beneficiaries consider depends largely on the identity of the insuring firm, the premium, the deductible, and whether the plan includes any form of supplemental coverage during the coverage gap. The impact of the firm and deductible alone are substantial. Shutting down the firm consideration effects more than triples the average consideration set size and improves plan choice by increasing the chance beneficiaries consider their optimal plan. The role of the firm in consideration is especially interesting and encourages future research. The strong effect of

firms on consideration might reflect prior insurance experience of beneficiaries or firm familiarity through social and spousal influences. It is also plausible that extensive firm advertising is at play. Insofar as firm advertising may be steering beneficiaries towards sub-optimal choices through limited consideration, is is important for policymakers to have an understanding of those effects.

Results additionally highlight the importance of accounting for consideration when estimating risk preferences. I find estimates of mean risk aversion more than twice that of a full consideration model. My results on preferences and consideration are informative to policymakers as Americans increasingly encounter choices over health insurance plans. With influence over product standardization, presentation of information, and firm behavior, policymakers may be able to harness the information about *how* beneficiaries are choosing prescription drug plans to help remove the obstacles that prevent so many beneficiaries from considering and choosing their optimal plan. It may be desirable to incorporate into the enrollment process questions eliciting risk preferences in a manner similar to the Health and Retirement Study to present plans in a manner that reflects which plans are likely to be optimal for an individual. In light of strong firm effects and the substantial reduction in the cost of limited consideration in the absence of firm effects, it may be advisable for CMS to evaluate policies on firm marketing, especially with regards to the recent regulatory changes on marketing materials and meaningful difference requirements. In any such intervention, a clearer understanding of how individuals behave in this market is of utmost importance. My model and estimates provide new insight into that process.

APPENDIX A

APPENDIX OF CHAPTER 2: HETEROGENEOUS CHOICE SETS AND PREFERENCES

A.1 Supplemental Materials

Table A.1: Auto Collision Claim Rate Regression - Part 1

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

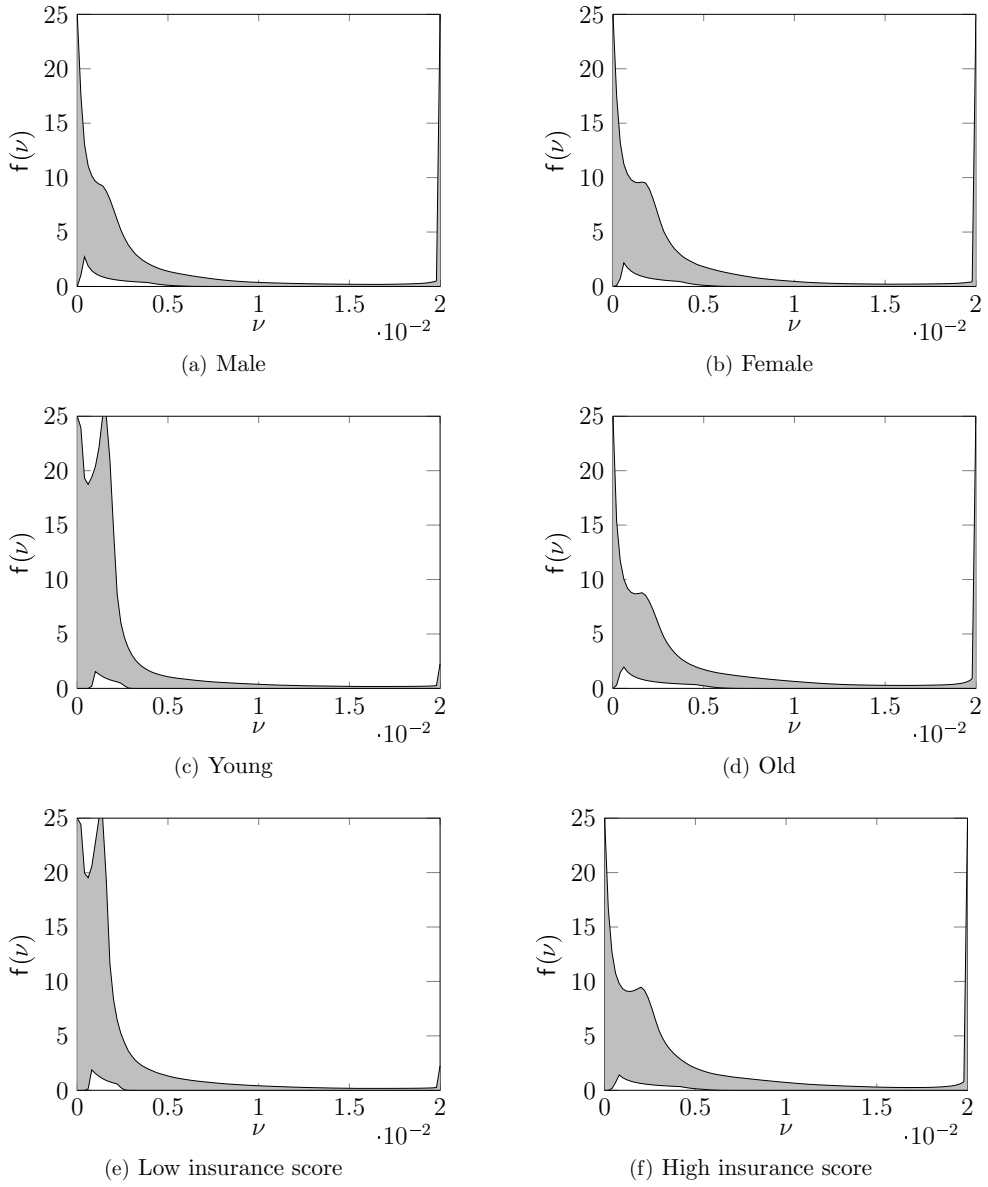
Note: Poisson panel regression with random effects (1,349,853 observations). Insurance score is a credit based risk score. Territory codes indicate rating territories, which are based on actuarial risk factors such as traffic and weather patterns, population demographics, wildlife density, and the cost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

Table continued on next page.

Table A.2: Auto Collision Claim Rate Regression - Part 2

Variable	Coefficient	Standard error
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: Table continued from previous page.



(a) Male (b) Female

(c) Young (d) Old

(e) Low insurance score (f) High insurance score

Figure A.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.

A.2 Theory

A.2.1 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 $\tilde{\boldsymbol{\nu}}_{\boldsymbol{i}}$ denote the first $|\mathcal{D}|$ components of $\boldsymbol{\nu_i}$, and let $\{\mathrm{e}_c : c \in \mathcal{D}\}$ be a collection of $|\mathcal{D}| \times 1$ standard basis vectors whose cth component equals one. Then $\epsilon_{ic} = \mathbf{e}_c^\top \tilde{\boldsymbol{\nu}}_{\boldsymbol{i}}$. For reasons we explain in Section 2.4.3, 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.2.2 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 sim-

plicity 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.2.1 (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.2.2 (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.2.1 (Artstein's Theorem): *A random vector y and a random set Y can be realized on the same probability space as random elements y' and Y', distributed as y and Y respectively, so that $P(y' \in Y') = 1$, if and only if*

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

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

LEMMA A.2.1: *The set $D^*_\kappa(\mathbf{x}_i, \boldsymbol{\nu}_i; \boldsymbol{\delta})$ in equation* (2.2) *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.2.1 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.2.3 Proof of Theorem 2.3.1

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 2.2.2(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 $c \in \mathcal{D}$

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

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 (2.6). 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 (2.6) 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 (2.6) holds. Conversely, take a value of $\boldsymbol{\theta}$ for which the inequalities in equation (2.6) are satisfied. Then, by Theorem A.2.1, 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.2.4 Sufficient Collection of Test Sets K

Theorem 2.3.1 and Corollary 2.3.1 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 2.3.1 also suffices for Theorem 2.3.2.

THEOREM A.2.2: *Let the assumptions of Theorem 2.3.1 hold. Then the following steps yield a sufficient collection of sets K, denoted $\mathbb{K}$, on which to check the inequalities in equation* (2.6) *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 $\mathbb{K} = \mathbb{K} \backslash K$;*[1]

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 2.4–2.5, 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

[1] 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.

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 (2.6) 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 (2.6) is additive, and therefore that inequality evaluated at K is implied by the ones evaluated at $\{c_j\}$ and at K_{q-1}. □

Depending on the structure of the realizations of the random set $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$, Theorem A.2.2 can be further simplified. The following corollary provides an example.

COROLLARY A.2.1: *Let Assumptions 2.2.1 and 2.2.2 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.2.2 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 $K \subset \mathcal{D}$, $\Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \cap K \neq \varnothing; \boldsymbol{\gamma}) = 1 - \Pr(D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta}) \subset K^C; \boldsymbol{\gamma})$, and therefore

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

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 2.3.1 is maintained, the logic of Theorem A.2.2 can be used to obtain a collection of sufficient test sets K on which to verify the inequalities in (2.8), 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 2.3.1, when Assumption 2.3.1 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.2.2 applied only to $D^*_\kappa(\mathbf{x}, \boldsymbol{\nu}; \boldsymbol{\delta})$ deliver the sufficient collection of sets K on which to verify (2.8) 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.2.2 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 (2.8) 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 (2.8) to obtain the sharp identification region for π_κ.

The two corollaries that follow illustrate the specific adaptations of Theorem A.2.2 that we use in our application in Sections 2.4–2.5. Proofs are omitted because the corollaries follow immediately from Theorem A.2.2.

COROLLARY A.2.2: *Let $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 3$. Suppose that all assumptions in Corollary 2.3.1 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* (2.8) *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.2.3: *Let $\mathcal{D} = \{c_1, c_2, c_3, c_4, c_5\}$ and $\kappa = 3$. Suppose that all assumptions in Corollary 2.3.1 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* (2.8) *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 2.4–2.5, the number of inequalities obtained through application of Theorem A.2.2 and Corollaries A.2.2–A.2.3 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.

A.3 Statistical Inference

The sample moments that we use to obtain the confidence intervals for (projections of) $\boldsymbol{\theta}$ in Section 2.5 are of the form:

$$\bar{m}_{n,K,j}(\boldsymbol{\theta}) = \widehat{\Pr}(d_i \in K | (\mu_i, \mathbf{p}_i) \in B_j) - \int_{B_j} P(D^*_\kappa(\mu, \mathbf{p}) \cap K \neq \varnothing; \boldsymbol{\gamma}) d\mu d\mathbf{p}, \quad \text{(A.4)}$$

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 (A.4) is computed using numerical approximation.

We obtain confidence regions based on the procedure proposed by Andrews and Soares (2010), as for example in Figure 2.3, and confidence intervals based on the procedure proposed by Kaido et al. (2016), as for example in Table 2.3. 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)):

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

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,

$$\liminf_{n \to \infty} \inf_{\mathsf{P} \in \mathcal{P}} \inf_{\boldsymbol{\theta} \in \Theta_I} \mathsf{P}(\boldsymbol{\theta} \in CS) \geqslant 1 - \alpha. \tag{A.5}$$

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.[2] 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

[2]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.

of $T_n(\boldsymbol{\theta})$. We use the hard-threshold GMS function proposed by Andrews and Soares (2010):[3]

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

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. (2016). 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. (2016, 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. (2016, 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 (A.5), is uniformly asymptotically covered with probability $1-\alpha$ over a large class of probability distributions $\mathcal{P}$ described in Kaido et al. (2016). Formally,

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

[3]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 2.2.2.

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 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. (2016, Section 2.2). We choose the value of ρ as follows. We begin with the recommendation in Kaido et al. (2016, 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.

APPENDIX B

APPENDIX OF CHAPTER 3: INSURANCE CHOICE WITH NON-MONETARY PLAN ATTRIBUTES: LIMITED CONSIDERATION IN MEDICARE PART D

B.1 Cost Calculator and Distribution Estimation

Using detailed data on the plan cost structure, I construct a program to compute the out-of-pocket costs for any series of ordered drug claims under every available plan in 2010 in the 34 major regions in the United States. The detailed claims data include information about gross and out-of-pocket realized drug costs under chosen plans, but a cost calculator is required to compute the counterfactual drug expenditures under the plans individuals did not select. The first step of the calculator is the collect the relevant set of plans to construct costs based on CMS region.

Prior to any calculation, I first assign every drug in the claims data, pricing data, and plan formulary a unique reference National Drug Code (NDC). Theoretically, each drug - defined by molecule, dosage, route of administration (tablet, injection, etc.), and brand name (if applicable) - is identified by a numeric code. The claims data identify drugs by NDC and a CMS created identifier called the Formulary RX ID. The plan formulary uses NDCs and another numeric system called RXCUIs. The base price data uses only NDCs to identify drugs. NDCs are not, however, unique identifiers.[1] The same drug may be listed under multiple NDCs within the data. Moreover, an individual's claim may record a drug under one NDC but the formulary for an available plan may use

[1]Formulary RX IDs and RXCUIs are similarly not unique.

an alternative NDC. A naive mapping could erroneously determine the drug is excluded from the formulary. To address this I create a mapping of NDCs based on Formulary RX IDs. For each NDC in the claims data, I collect all Formulary RX IDs ever attributed to it. I then take the set of Formulary RX IDs and collect all of the NDCs to which those identifiers are ever linked. I repeat that process one additional time, and the resulting set of NDCs are deemed to represent the same drug. I then assign all linked NDCs the same unique numeric identifier in the claims, pricing, and formulary data.

B.1.1 Drug Cost Calculator

For any sequence of claims, I identify the coverage classification of each drug under each available plan. For every drug included in a plan's formulary, I determine the tier of coverage and whether that tier is covered in the donut hole. I also determine the base price of the claim by scaling the negotiated price of a 30 days supply of each drug under each plan to correspond to the days supply claimed. My calculator then processes the claims sequentially, determining the coverage phase and applying the relevant cost-sharing based on tier and phase. In the event a claim straddles multiple coverage phases, I prorate the claim across spending zones in the manner CMS does in practice. The calculator keeps a running total of gross and out-of-pocket spending throughout the series of claims.

To assess the performance of the cost calculator, I compare the estimated out-of-pocket spending for each beneficiary's chosen plan to their realized out-of-pocket costs in the claims data. When I use as the base price of a drug the gross

cost listed in the claims data, predicted and observed out-of-pocket spending have a correlation in excess of .95 for individuals across regions. In practice, I use the negotiated base prices listed in the pricing data to account for differences in base prices across plans. Occasionally there is a discrepancy between the information in the pricing file and what is reported in the claims data. Once I incorporate the negotiated base prices, the correlation between predicted and observed spending is .93 among the analysis sample. This simple test of accuracy is assuring, especially as I made a number of small simplifications in constructing the calculator that would prevent perfect prediction. I treated all claims as filed through in-network pharmacies and pro-rated one month cost-sharing for tractability.[2] In the catastrophic coverage phase, I treat all claims as though they are branded drugs. Out-of-pocket costs for those claims are therefore computed as the maximum between a 5% coinsurance and a $6.30 copay. In practice, for generic drugs, the beneficiary pays the maximum of a 5% coinsurance and a $2.50 copay. The data I use does not include information on whether a drug is branded or generic. However, few individuals enter the catastrophic coverage phase at all, and the differences in cost between these two pricing schemes is small.

B.1.2 Distribution of Out-of-Pocket Costs and Variance Estimation

Ad described in Section 3.4.2, higher order moments of the distribution of drug costs an individual expects under different plans requires an approximation of

[2] In general, the three month copay was simply 3 times the one month copay, making this simplification quite innocuous.

the distribution of out-of-pocket drug costs under each available plan. In practice, it is the variance of costs for which I need estimates. To this end, I assign each individual in my sample into a bin of "similar individuals" based on their average monthly gross drug costs and average monthly "effective" claim counts during their 2010 tenure. Effective claim counts adjusts counts for the number of months a claim covers. For example, if a beneficiary filled a claim for a 90 day supply, it is treated as effectively 3 claims. Average claim counts are are classified as one of the following: between 0 and 1, between 1 and 2, between 2 and 3, between 3 and 4, between 4 and 10, and more than 10. These claims bins are crossed with quintiles of average monthly gross spend. An additional bin of individuals with zero claims and zero spend is also defined. Bins with fewer than 100 individuals are dropped.

To estimate the cost distribution within each bin, I construct a sample of individuals without ESRD who are enrolled in a Part D plan for some portion of 2010 and the entirety of 2011. I use their 2010 claims experience to categorize them into one of the bins described above. I then randomly select 100 individuals from each bin and pass their entire 2011 claims experiences through the cost calculator for every plan. I compile monthly running totals of out-of-pocket spend for each randomly sampled individual. To adjust for the evolution of drug expenditure over time, I deflate all 2011 costs by the average ratio of 2010 spend compared to 2011. For every individual in my analysis sample, denoting their months of 2010 coverage by m, the variance of out-of-pocket costs in each plan is computed as the variance of the random sample's deflated out-of-pocket costs for m months of 2011.

B.2 Estimation

B.2.1 Consideration Probabilities

In Section 3.6.3 I describe the intuition behind the consideration probabilities, φ_j. In practice, I model each plan's consideration probability, φ_j, as a function of plan j's characteristics listed above:

$$\varphi_j = f(\text{firm}_j, \text{premium}_j, \text{deductible}_j, \text{gap}_j, \text{Top100}_j, \text{AvgCS}_j)$$

To ensure consideration probabilities are in the unit interval, I impose the following functional form:

$$\varphi_j = \phi_{firm_j}\phi_{prem_j}\phi_{ded_j}\phi_{gap_j}\phi_{top100_j}\phi_{AvgCS_j},$$

where $\phi_{firm_j} \in [0, 1]$ is the base consideration probability of the firm offering plan j, constant across all plans offered by that firm in the California market. The plan attributes enter consideration multiplicatively as well, with all δ terms $\in [0, 1]$

$$\begin{aligned}
\phi_{prem_j} &= \delta_{prem}^{PremRatio}, \\
\phi_{ded} &= \delta_{ded}^{DedRatio} \\
\phi_{Gap} &= \begin{cases} \delta_{gap} \text{ if No Gap} \\ 1 \text{ if Gap}, \end{cases} \\
\phi_{Top100} &= \delta_{top100}^{(max(top100)-top100_j)}, \\
\phi_{AvgCS} &= \delta_{avgcs}^{(AvgCS_j - min(AvgCS))}.
\end{aligned}$$

Both ϕ_{prem_j} and ϕ_{ded_j}, which govern the roles of premium and deductible, respectively, depend on the ratio of a plan's premium and deductible relative

to the maximum in the market. I define $PremRatio_j \equiv \frac{Prem_j - min(Prem)}{max(Prem) - min(Prem)}$ and $DedRatio_j \equiv \frac{Deduc_j}{max(Deduc)}$.

B.3 Robustness Analysis

In my baseline analysis I assume perfect foresight of expected out-of-pocket drug costs. As a robustness check, I estimate my model using an alternative specification of expected out-of-pocket costs. Table B.1 presents estimates of the consideration impact of plan attributes. The reduction in consideration that occurs between the best and worst premiums and deductibles is larger in this specification. All else equal, these estimates suggest the highest premium plan is considered only 7% as much as the lowest premium plan. Plans with $310 deductibles receive 12% as much consideration as equivalent $0 plans. The impact of gap coverage is similar but slightly milder. Similar to the baseline analysis, the count of top 100 drugs covered and the average cost-share of a plan do not impact consideration. Figure B.1 plots firm base consideration probabilities in the same manner as Figure 3.5. The same patterns emerge as in the baseline results. Estimates of risk aversion are slightly higher than in the baseline analysis, but on the whole similar. Table B.2 presents estimates and confidence intervals for the mean and variance of risk aversion. There is substantial overlap in the confidence intervals for both statistics. These estimates show that the results in the baseline analysis are not driven by the assumption of perfect foresight of expected out-of-pocket expenses.

Table B.1: Robustness Results: Consideration Impact of Plan Attributes

	Estimate	95% CI
δ_{prem}	0.072	[0.051, 0.104]
δ_{ded}	0.120	[0.106, 0.136]
δ_{gap}	0.906	[0.822, 1.000]
δ_{top100}	1.000	[1.000, 1.000]
δ_{avgcs}	1.000	[1.000, 1.000]

Notes: All δ terms are defined between 0 and 1 and reflect how much consideration a plan with the worst value of an attribute is considered relative to an equivalent plan with the best value of the attribute. Confidence intervals based on 1,000 bootstraps with sub-sampling to correct for estimates on the boundary.

Table B.2: Robustness Estimates: Risk Preferences

	Estimate	Risk Premium	95% CI
Limited Consideration			
$\mathbb{E}$(Risk Aversion)	$1.08 \cdot 10^{-3}$	\$117	$[5.31 \cdot 10^{-4}, 1.70 \cdot 10^{-3}]$
Var(Risk Aversion)	$3.99 \cdot 10^{-6}$		$[9.53 \cdot 10^{-7}, 8.37 \cdot 10^{-6}]$

Notes: CI based on 1,000 bootstraps. In limited consideration model, sub-sampling used to correct for estimates on the boundary. Risk premium is calculated for a beneficiary facing a lottery that results in a loss of \$1,000 with 25% probability.

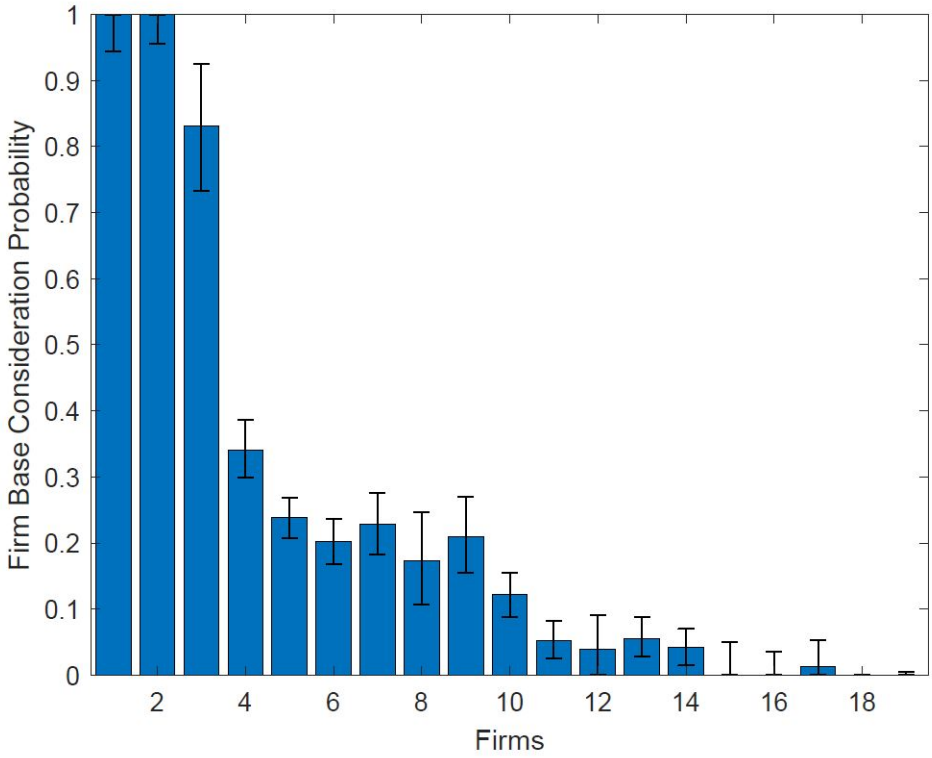


Figure B.1: Robustness Results: Firm Base Consideration Probabilities

Notes: Firms are ordered based on estimated base consideration probabilities in the baseline model, as in Figure 3.5. Error bars present 95% confidence intervals based on 1,000 bootstrap repetitions with sub-sampling to adjust for estimates on the boundary.

BIBLIOGRAPHY

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.

ABALUCK, J. AND J. GRUBER (2011): "Choice inconsistencies among the elderly: evidence from plan choice in the Medicare Part D program," *American Economic Review*, 101, 1180–1210.

——— (2016a): "Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program: Reply," *American Economic Review*, 106, 3962–87.

——— (2016b): "Evolving choice inconsistencies in choice of prescription drug insurance," *American Economic Review*, 106, 2145–84.

ANDREWS, D. W. 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., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2019a): "Heterogeneous choice sets and preferences," *arXiv preprint arXiv:1907.02337*.

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 (2019b): "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 option," *The Quarterly Journal of Economics*, 132, 1319–1372.

BROWN, Z. Y. AND J. JEON (2019): "Endogenous Information Acquisition and Insurance Choice," .

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 (2012): "From Aggregate Betting Data to Individual Risk Preferences," Discussion Paper 1213-08, Department of Economics, Columbia University.

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.

DIEBOLD, J. (2016): "The effects of medicare part D on health outcomes of newly covered medicare beneficiaries," *The Journals of Gerontology: Series B*, 73, 890–900.

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.

DURANTE, A. AND L. CHEN (2019): "Report on the economic well-being of US households in 2018 [Report for the Board of Governors of the Federal Reserve System]. Washington, DC: Board of Governors of the Federal Reserve System," .

ERICSON, K. M. AND A. STARC (2012): "Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts connector," *American Economic Review*, 102, 493–97.

ERICSON, K. M. M. AND A. STARC (2016): "How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange," *Journal of Health Economics*, 50, 71–85.

FOSGERAU, M., E. MELO, A. D. PALMA, AND M. SHUM (2017): "Discrete choice and rational inattention: A general equivalence result," *Available at SSRN 2889048*.

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 US 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.

HANDEL, B. R. AND J. T. KOLSTAD (2015): "Health insurance for" humans": Information frictions, plan choice, and consumer welfare," *American Economic Review*, 105, 2449–2500.

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., A. LEIVE, D. MCFADDEN, AND J. WINTER (2013): "Plan selection in Medicare Part D: Evidence from administrative data," *Journal of Health Economics*, 32, 1325–1344.

HEISS, F., D. MCFADDEN, AND J. WINTER (2010): "Mind the gap! Consumer perceptions and choices of Medicare Part D prescription drug plans," in *Research findings in the economics of aging*, University of Chicago Press, 413–481.

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.

HO, K., J. HOGAN, AND F. SCOTT MORTON (2017): "The impact of consumer inattention on insurer pricing in the Medicare Part D program," *The RAND Journal of Economics*, 48, 877–905.

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

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

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 (2016): "Confidence intervals for projections of partially identified parameters," *arXiv preprint arXiv:1601.00934*.

KAISER FAMILY FOUNDATION AND THE HARVARD SCHOOL OF PUBLIC HEALTH (2006): *Seniors and the Medicare Prescription Drug Benefit*, https://www.kff.org/medicare/poll-finding/seniors-and-the-medicare-prescription-drug-benefit/.

KEANE, M. P., J. D. KETCHAM, N. V. KUMINOFF, AND T. NEAL (2019): "Evaluating Consumers' Choices of Medicare Part D Plans: A Study in Behavioral Welfare Economics," Tech. rep., National Bureau of Economic Research.

KESTERNICH, I., F. HEISS, D. MCFADDEN, AND J. WINTER (2013): "Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D," *Journal of Health Economics*, 32, 1313–1324.

KETCHAM, J. D., N. V. KUMINOFF, AND C. A. POWERS (2016): "Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program: Comment," *American Economic Review*, 106, 3932–61.

——— (2019): "Estimating the Heterogeneous Welfare Effects of Choice Architecture," *International Economic Review*.

KETCHAM, J. D., C. LUCARELLI, E. J. MIRAVETE, AND M. C. ROEBUCK (2012): "Sinking, swimming, or learning to swim in Medicare Part D," *American Economic Review*, 102, 2639–73.

KETCHAM, J. D., C. LUCARELLI, AND C. A. POWERS (2015): "Paying attention or paying too much in Medicare Part D," *American Economic Review*, 105, 204–33.

KIM, J. B., P. ALBUQUERQUE, AND B. J. BRONNENBERG (2010): "Online Demand Under Limited Consumer Search," *Marketing Science*, 29, 1001–1023.

KLING, J. R., S. MULLAINATHAN, E. SHAFIR, L. C. VERMEULEN, AND M. V. WROBEL (2012): "Comparison friction: Experimental evidence from Medicare drug plans," *The Quarterly Journal of Economics*, 127, 199–235.

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.

LUCARELLI, C., J. PRINCE, AND K. SIMON (2012): "The welfare impact of reducing choice in Medicare Part D: A comparison of two regulation strategies," *International Economic Review*, 53, 1155–1177.

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.

MATĚJKA, F. AND A. MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–98.

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. (1974): "Frontiers in Econometrics, chapter Conditional logit analysis of qualitative choice behavior," .

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

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

MOLCHANOV, I. AND F. MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge: Cambridge University Press.

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

POLYAKOVA, M. (2016): "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 8, 165–95.

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.

SEMILLA, A. P., F. CHEN, AND T. M. DALL (2015): "Reductions in mortality among Medicare beneficiaries following the implementation of Medicare Part D." *The American journal of managed care*, 21, s165–71.

SIMON, H. A. (1959): "Theories of decision-making in economics and behavioral science," *The American economic review*, 49, 253–283.

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.

TVERSKY, A. (1972): "Elimination by aspects: A theory of choice." *Psychological review*, 79, 281.

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.