Estimating Multinomial Choice Models with Unobserved Choice Sets

Zhentong Lu*
Shanghai University of Finance and Economics
lu.zhentong@shufe.edu.cn
zhentonglu.com

February 14, 2017

Abstract

This paper proposes a new approach to estimating multinomial choice models when each consumer's actual choice set is unobservable but could be bounded by two known sets, i.e., the largest and smallest possible choice sets. The bounds on choice set, combined with a monotonicity property derived from utility maximization, imply a system of inequality restrictions on observed choice probabilities that could be used to identify and estimate the model. A key insight is that the identification of random utility model could be achieved without exact information on consumers' choice sets, which strictly generalizes the identification result of the standard multinomial choice model. The effectiveness of the proposed approach is demonstrated via a range of Monte Carlo experiments as well as an empirical application to consumer demand for potato chips using household scanner data.

Keywords: Discrete Choice; Choice Set Heterogeneity; Moment Inequalities; Scanner Data
JEL: C14, C50, L00, M30.

1 Introduction

Understanding consumer preference from revealed choices is a basic issue in economics. Originated from Marschak (1960) and subsequently developed by McFadden (1973), Berry, Levinsohn, and Pakes (1995), etc., the random utility model has become the standard econometric tool for eliciting preference using data on consumer/product characteristics and purchase decisions. The central premise of the random utility model is utility maximization, i.e., a consumer chooses the product that gives her/him the highest utility from the choice set. Both utility function and choice set

*This paper is based on the first chapter of my PhD dissertation at Wisconsin-Madison. Earlier version was circulated under the title "A Moment Inequality Approach to Estimating Multinomial Choice Models with Unobserved Consideration Sets". I am indebted to my advisors Amit Gandhi and Xiaoxia Shi, and other committee members, Ken Hendricks and Alan Sorensen, for their invaluable guidance. Also, I thank the editor Aviv Nevo and two anonymous referees for their important comments and advice that have greatly improved the paper. Finally, I am grateful to many individuals for their helpful feedback on this project. All remaining errors are mine.

are essential in determining a consumer's final choice. The traditional literature on discrete choice models (see Train (2009) for a comprehensive coverage on the topic) focuses on modeling the utility function (preference) and treats a consumer's choice set as exogenously given; in practice, to estimate the model, empirical researchers have to impute choice sets for consumers, which usually include all the products available in a given market.

But the imputed set may not be the underlying true choice set of a consumer. As a basic consensus of the large literature in marketing and psychology concerning consideration set,[1] see, e.g., Gilbride and Allenby (2006), Gilbride and Allenby (2004), Hauser and Wernerfelt (1990), Hauser, Toubia, Evgeniou, Befurt, and Dzyabura (2010), Jedidi and Kohli (2005), Liu and Arora (2011), Roberts and Lattin (1991) and Shocker, Ben-Akiva, Boccara, and Nedungadi (1991), consumers' actual choice sets are endogenously chosen and typically small comparing to the whole set of products available in a market due to cognitive capacity limitations. Both theoretical analysis, see, e.g., Masatlioglu, Nakajima, and Ozbay (2012); Manzini and Mariotti (2014); Eliaz and Spiegler (2011), and empirical evidence, see, e.g., Goeree (2008), Draganska and Klapper (2010), suggest that ignoring the choice set heterogeneity has important consequences for learning preference from consumer choice data.

To address this concern, we could extend the random utility framework in a way that both choice set formation and preference are jointly modeled. However, the fundamental challenge of modeling choice set formation is that the number of possible choice sets, which grows exponentially with the number of products in a market, is impractically large in many empirical applications. To circumvent this difficulty, the existing literature imposes detailed parametric structures and identification assumptions on choice set formation, e.g., Ben-Akiva and Boccara (1995), Chiang, Chib, and Narasimhan (1999), Goeree (2008), Bruno and Vilcassim (2008), Gentry (2011), Conlon and Mortimer (2008), Mehta, Rajiv, and Srinivasan (2003), Hortacsu and Syverson (2004), Santos, Hortacsu, and Wildenbeest (2012), Van Nierop, Bronnenberg, Paap, Wedel, and Franses (2010), Paola and Marco (2013) and Pires (2012). These different specifications of choice set formation are plausible in certain applications. But this is not always the case - these structures and assumptions could be restrictive for some empirical applications and thus increase the risk of misspecification. Furthermore, even with a well specified model of choice set formation, researchers have to use simulation and sampling techniques to integrate over all the possible choice sets when estimating the model, see e.g., Chiang, Chib, and Narasimhan (1999), Goeree (2008),[2] Moraga-González, Sándor, and Wildenbeest (2009) and Bruno and Vilcassim (2008), which can be computationally demanding for many empirical applications.

[1]The concept of consideration set arises in marketing literature and it is subtly different from choice set in standard discrete choice models. In this paper, I respect this discrepancy and use "choice set" to refer the set of products over which a consumer maximizes utility. More precisely, if the random utility has full support, e.g., logit model, then the products in the choice set have strictly positive choice probabilities while the products outside the choice set have zero choice probability. I thank both referees for suggesting this clarification.

[2]Exploiting the independence between preference and choice set formation, Goeree (2008) simulates one choice set for each consumer to reduce the computational burden. In the Monte Carlo section, I include this estimation strategy as a comparison with the proposed approach of this paper, both in terms of estimation accuracy and computational time. I thank one referee for the comments on this issue.

This paper develops an alternative approach to getting around the dimensionality issue while placing limited and intuitive restrictions on the choice set generating process. The main contribution is to show that, when consumers' choice sets are unobservable (to econometricians) but could be informatively bounded, we could still identify and estimate the random utility model based on a system of conditional moment inequalities without modeling the choice set formation. The moment inequalities are effectively constructed from a pair predicted bounds on the observed choice probability for each consumer/product, which are in turn based on two elements: (1) bounds (largest and smallest possible choice sets) on each consumer's actual choice set; (2) monotonicity of choice function with respect to choice set: if an alternative x is chosen from a set T, and x is also an element of a subset $S \subseteq T$, then the x must be chosen from S, which is a immediate implication of utility maximization assumption and also known as Chernoff's condition (Chernoff, 1954) or Sen's property α (Sen, 1971) in choice theory.

Comparing with the common strategy of modeling choice set formation, my approach essentially trades parametric functional form assumptions for nonparametric support restrictions on the underlying choice set distribution. An important advantage of this approach is that it only requires two sets (the largest and smallest) to construct bounds on the choice probability for each consumer/product, and is thus conceptually and computationally easy to implement - there is no need to integrate over the distribution of unobserved choice set in the estimation process as typically required with a fully specified model of choice set formation.

The moment inequalities could be informative enough to identify the random utility model. And the key requirement is that, for a positive measure of consumers, the true choice set is close enough to either side (not both sides) of the bounds, so that the associated inequalities are violated when the parameter vector deviates from its true value. Also, using a criterion measuring violations of the inequalities, a simple analog estimator is defined and shown to be consistent under regularity conditions.

I perform an extensive set of Monte Carlo experiments to evaluate the bounds approach under various scenarios and compare it with alternative estimation strategies, which either assume fixed choice sets or impose choice set formation models. The results provide support to the theory in the sense that, as long as the bounds are correctly specified and identification condition is met, the bounds approach works well in terms of yielding accurate point estimates and correcting the biases caused by misspecified choice sets. Also, bounds approach is computationally lighter than alternative methods that incorporate choice set formation, e.g., independent sampling model (Goeree (2008)), and sequential search model (Hortacsu and Syverson (2004)).

Finally, the bounds approach is applied to demand estimation using the IRI household panel and store scanner data, which is a leading source for demand information of consumer packaged goods. Specifically, I focus on potato chips category and define the bounds on the choice set for each purchase occasion as follows. The largest possible choice set includes all the products available on the shelf at the store/week where the purchase happened; the smallest one is comprised of those ever purchased by the household in an extended period of time prior to the current purchase.

Comparing with the standard method that assumes a fixed choice set for each purchase occasion, which could be misspecified, the bounds approach generates substantially different patterns of consumer demand, notably more consumer heterogeneity in the responses to the key marketing mix variables, including price, display and feature advertising.

The rest of the paper is organized as follows. Section 2 lays out the basic setup. Section 3 derives the bounds and moment inequalities, introduces the sufficient conditions for identification and presents a consistent point estimator based on the conditional moment inequalities. Sections 4 and 5 report the results of Monte Carlo experiments and the empirical applications. Section 6 concludes. All technical proofs are in the Appendix.

2 Model

2.1 Setup

Consider a generic market consisting of a set of differentiated products $\mathcal{J} = \{0, 1, ..., J\}$ and a population of independent and ex ante identical consumers, denoted by $\mathbf{I}$. The product labeled by 0 is the "outside option" and those labeled by $j = 1, \ldots, J$ are the "inside goods". Each consumer $i \in \mathbf{I}$ is associated with a matrix of consumer-/product-specific observables $X_i \equiv (X_{i1}, ..., X_{iJ})$, where $X_{ij} \in \mathbf{R}^K$ typically consists of product characteristics, e.g., price, consumer attributes, e.g., income, as well as the interactions between them. Each consumer i's purchase decision is observable and designated as $d_i \equiv (d_{i0}, d_{i1} ..., d_{iJ})$, where $d_{ij} = 1$ if consumer i chooses product j.

Consumer preference is represented by a linear random utility model;[3] the indirect utility to consumer i from choosing j is

$$u_{ij} = X_{ij}\beta + \varepsilon_{ij}, \tag{2.1}$$

where

$$\varepsilon_i \equiv (\varepsilon_{i0}, ..., \varepsilon_{iJ}) \sim F(\cdot | X_i; \lambda)$$

represents an idiosyncratic preference shock, β and λ are parameters of interests. Allowing X to enter F_ε, the specification nests the usual random coefficients logit model (see, e.g., Berry, Levinsohn, and Pakes, 2004),

$$\begin{aligned} u_{ij} &= X_{ij}(\beta + \upsilon_i) + \epsilon_{ij} \\ &= X_{ij}\beta + \underbrace{(X_{ij}\upsilon_i + \epsilon_{ij})}_{\varepsilon_{ij}}, \end{aligned} \tag{2.2}$$

where the random coefficient $\upsilon_i \sim N(0, \Sigma(\lambda))$, ϵ_{ij} follows the standard Gumbel distribution and is i.i.d. across i, j.

Each consumer i chooses a product that maximizes utility from his/her choice set C_i, which is unobservable to econometricians. The unobserved choice set frustrates us from computing consumer

[3]Assuming linearity is only for expositional convenience and the method proposed in this paper applies to general random utility models.

i's choice probability of product j,

$$\sigma_j(X_i, C_i; \theta) = \begin{cases} \int 1\left(u_{ij} = \max_{k \in C_i}\{u_{ik}\}\right) dF\left(\varepsilon_i \,|X_i; \lambda\right) & \text{if } j \in C_i \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta \equiv (\beta, \lambda)$ represents the parameters in the random utility model collectively.

The standard approach to proceed is to treat C_i as unobserved heterogeneity and specify a model of choice set formation that generates a distribution of C_i,

$$\left\{\Pr\left(C_i = \mathcal{I}\right) : \mathcal{I} \in 2^{\mathcal{J}}\right\}, \tag{2.3}$$

which is then embedded into the random utility model to predict choice probabilities, i.e.,

$$\Pr\left(d_{ij} = 1 \middle| X_i\right) = \int \sum_{\mathcal{I} \in 2^{\mathcal{J}}} 1\left(u_{ij} = \max_{k \in C_i}\{u_{ik}\}\right) \Pr\left(C_i = \mathcal{I}\right) dF\left(\varepsilon_i \,|X_i; \lambda\right).$$

Matching the predicted choice probabilities with observed ones provides the foundation for jointly estimating preference and choice set formation. In fact, many models of choice set formation, either random sampling or optimal searching, could be regarded as different specifications and restrictions imposed on (2.3), see Goeree (2008), Hortacsu and Syverson (2004), Santos, Hortacsu, and Wildenbeest (2012), etc.

2.2 Bounds on Choice Set

Instead of modeling choice set formation, this paper proposes an alternative solution to the problem of unobservable choice sets by directly restricting the support of distribution (2.3), which is formally stated in the following assumption.

Assumption 1. *Each consumer i's choice set C_i is bounded by two sets C_i^{sup} and C_i^{sub} such that $C_i^{sub} \subseteq C_i \subseteq C_i^{sup} \subseteq \mathcal{J}$.*

An immediate implication of Assumption 1 is that, $\Pr\left(C_i = \mathcal{I}\right) = 0$ if $\mathcal{I} \subset C_i^{sub}$ or $C_i^{sup} \subset \mathcal{I}$, i.e., $\mathcal{I}$ does not lie within the bounds. The major motivation for Assumption 1 is that in most empirical applications of demand estimation, data could not provide perfect information on choice set[4] but might be informative enough for defining bounds on it. For instance, in the empirical application of this paper, the upper bound C_i^{sup} includes all the products available in a given market, which is often treated as consumer i's true choice set in the standard approach; the lower bound C_i^{sub} is defined as the set of products previously bought (up to a certain length of time period) by the consumer, since each consumer's purchase history is observable[5].

[4] An exception is Draganska and Klapper (2010) who uses survey data on choice sets.

[5] A related research, Chiang, Chib, and Narasimhan (1999), exploits consumer purchase history as the basis for modeling choice set heterogeneity.

Another example is the stock-out problem studied by Conlon and Mortimer (2008) and Musalem, Olivares, Bradlow, Terwiesch, and Corsten (2010), in which econometricians observe the product availability at the beginning and the end of a time interval, as well as consumer i's purchase during this time period. In this case, consumer i's choice set is unobservable because researchers do not know the product availability at the time that consumer i makes the purchase, however, the choice set could be safely assumed to be a subset of products available at the beginning (C_i^{sup}) and a superset at the end (C_i^{sub}) of the time interval.

In general, the bounds in any specific empirical application should be chosen on a case-by-case basis. Since the main focus of this paper is developing a method to exploit the bounds, I shall first present how to use them for identification/estimation, and postpone the discussions on choosing them to Subsection 3.4 [6].

Remark 1. A recent paper by Crawford, Griffith, and Iaria (2016) proposes using "sufficient set", which is assumed to be a subset of a consumer's unobserved true choice set, to estimate logit class models with unobserved choice set heterogeneity (see also McFadden (1978)). They provide several examples of constructing sufficient set by exploiting the structure of panel data, e.g., using consumer purchase history to define sufficient set. The notion of sufficient set is similar to the lower bound C_i^{sub} in Assumption 1, however, the two estimation strategies are fundamentally different and each has its own strengths and weaknesses. So the approach developed in this paper should be regarded as a complement to their method when researchers are dealing with the problem of unobserved choice sets.

3 Identification and Estimation

3.1 Bounds on Choice Probabilities

Given the well-defined random utility model (2.1), I now introduce the key theorem of this paper, which shows that bounds on choice set (Assumption 1) could be transformed into bounds on choice probabilities. The key to the transformation is an important monotonicity property implied by utility maximization, i.e., [7]

$$1\left(u_{ij} = \max_{k\in\mathcal{I}'}\{u_{ik}\}\right) \leq 1\left(u_{ij} = \max_{k\in C_i}\{u_{ik}\}\right) \leq 1\left(u_{ij} = \max_{k\in\mathcal{I}}\{u_{ik}\}\right), \tag{3.1}$$

for any $\mathcal{I}$ and $\mathcal{I}'$ such that $j \in \mathcal{I} \subseteq C_i \subseteq \mathcal{I}'$.

Theorem 1. *Suppose Assumption 1 holds. Then for each consumer i and product $j \in \mathcal{J}$,*

$$\underline{P}_j\left(X_i;\theta_0\right) \leq \Pr\left(d_{ij} = 1 | X_i\right) \leq \overline{P}_j\left(X_i;\theta_0\right), \tag{3.2}$$

[6] I thank the editor for the advice on adding such a discussion.

[7] This is also known as Chernoff's condition (Chernoff, 1954) or Sen's property α (Sen, 1971): if an alternative x is chosen from a set T, and x is also an element of a subset S of T, then the x must be chosen from S.

where

$$\underline{P}_j(X_i;\theta_0) \equiv E\left[1\left\{j \in C_i^{sub}\right\}\sigma_j\left(X_i, C_i^{sup};\theta_0\right)\middle| X_i\right],$$

$$\overline{P}_j(X_i;\theta_0) \equiv E\left[1\left\{j \in C_i^{sup}\right\}\sigma_j\left(X_i, C_i^{sub}\cup\{j\};\theta_0\right)\middle| X_i\right],$$

and θ_0 is the true value.

Proof. Observe that for any consumer i and each $j \in \mathcal{J}$, Assumption 1 and the property (3.1) implies that

$$\begin{aligned} 1\left\{j \in C_i^{sub}\right\}1\left(u_{ij} = \max_{k\in C_i^{sup}}\{u_{ik}\}\right) &\leq 1\left(u_{ij} = \max_{k\in C_i}\{u_{ik}\}\right) \\ &\leq 1\left\{j \in C_i^{sup}\right\}1\left(u_{ij} = \max_{k\in C_i^{sub}\cup\{j\}}\{u_{ik}\}\right), \end{aligned}$$

where the indicators $1\left\{j \in C_i^{sub}\right\}$ and $1\left\{j \in C_i^{sup}\right\}$ take care of the cases $j \in C_i^{sup}\setminus C_i^{sub}$ and $j \in \mathcal{J}\setminus C_i^{sup}$, respectively[8]. Now integrate the above inequalities over all possible C_i and we could obtain

$$\begin{aligned} &1\left\{j \in C_i^{sub}\right\}1\left(u_{ij} = \max_{k\in C_i^{sup}}\{u_{ik}\}\right)\sum_{\mathcal{I}\in 2^{\mathcal{J}}}\Pr\left(C_i = \mathcal{I}\right) \\ \leq &\sum_{\mathcal{I}\in 2^{\mathcal{J}}} 1\left(u_{ij} = \max_{k\in\mathcal{I}}\{u_{ik}\}\right)\Pr\left(C_i = \mathcal{I}\right) \\ \leq &1\left\{j \in C_i^{sup}\right\}1\left(u_{ij} = \max_{k\in C_i^{sub}\cup\{j\}}\{u_{ik}\}\right)\sum_{\mathcal{I}\in 2^{\mathcal{J}}}\Pr\left(C_i = \mathcal{I}\right). \end{aligned}$$

Next, note that $\sum_{\mathcal{I}\in 2^{\mathcal{J}}}\Pr\left(C_i = \mathcal{I}\right) = \Pr\left(j \in C_i\right)$ is the probability of product j being in C_i and satisfies

$$\Pr\left(j \in C_i\right)\begin{cases} = 1 & \text{if } j \in C_i^{sub} \\ \in [0,1] & \text{otherwise.}\end{cases}$$

Hence, we have

$$\begin{aligned} &1\left\{j \in C_i^{sub}\right\}1\left(u_{ij} = \max_{k\in C_i^{sup}}\{u_{ik}\}\right) \\ \leq &\sum_{\mathcal{I}\in 2^{\mathcal{J}}} 1\left(u_{ij} = \max_{k\in\mathcal{I}}\{u_{ik}\}\right)\Pr\left(C_i = \mathcal{I}\right) \\ \leq &1\left\{j \in C_i^{sup}\right\}1\left(u_{ij} = \max_{k\in C_i^{sub}\cup\{j\}}\{u_{ik}\}\right). \end{aligned} \tag{3.3}$$

[8]To see the necessity of indicator $1\left\{j \in C_i^{sub}\right\}$, note that $1\left(u_{ij} = \max_{k\in C_i}\{u_{ik}\}\right)$ might be zero for $j \in C_i^{sup}\setminus C_i^{sub}$ so we could only impose a conservative lower bound for this case. Similar logic could be applied to explain to the indicator $1\left\{j \in C_i^{sup}\right\}$ in the upper bound.

Finally, we obtain (3.2) by taking conditional expectation on (3.3) and applying law of iterated expectation. □

Theorem 1 states that for any given X_i, 1) for any $j \in C_i^{sub}$, the choice probability $\Pr\left(d_{ij}=1|\, X_i\right)$ is bounded below (above) by the counterfactual choice probability predicted by the random utility model when C_i^{sup} ($C_i^{sub} \cup \{j\}$) is the actual choice set of consumer i; 2) for any product that are in C_i^{sup} but not in C_i^{sub}, the lower bound on the choice probability is non-informative, i.e., 0, and the upper bound is computed at $C_i^{sub} \cup \{j\}$; 3) any product outside C_i^{sup} has zero bounds (and zero choice probability by construction).

Theorem 1 circumvents the dimensionality issue of modeling the choice set distribution because only two sets, i.e., C_i^{sub} and C_i^{sup}, are needed for constructing the inequality constraints. Also, as long as credible bounds on the choice set are available, we are free from imposing a model of choice set formation (could be arbitrary), which brings some robustness into the model. Certainly, a potential concern is that (3.2) may not be very informative about parameters, however, Subsection 3.2 shows that under certain conditions, the inequality (3.2) has substantial identification power and could point identify the parameters in the model.

Remark 2. Observe that Theorem 1 only relies on the assumption that bounds are valid (Assumption 1), regardless of how the bounds are constructed. In particular, suppose the lower bound C_i^{sub} is defined in a completely "endogenous" manner, e.g., $C_i^{sub} = \{K_i^*\}$, where K_i^* is realized chosen product, i.e., choice-based sampling [9], then the associated upper bounds on choice probabilities in (3.2) are still valid. However, such a bound will have little identification power, we will revisit this choice-based sampling issue after presenting the identification theorem in Subsection 3.2.

3.2 Point Identification via Moment Inequalities

In this subsection, I characterize the point identification condition following the similar arguments of Powell (1984) and Kahn and Tamer (2009). A practical advantage of focusing on point identification case is that, once it could be justified, we could obtain point estimate using standard numerical optimization techniques and thus avoid finding the set estimate, which can be challenging even with a small number of parameters.

In the context of moment inequalities, the point identification condition effectively requires that for any $\theta \neq \theta_0$, some inequalities are violated. Now let us define the following quantities to ease the interpretation of the identification condition. The first two measure the changes in bounds when the parameter θ deviates from its true value, i.e.,

$$R_{ij}^{sup}\left(\theta,\theta_0\right) = \underline{P}_j\left(X_i;\theta\right) - \underline{P}_j\left(X_i;\theta_0\right)$$

$$R_{ij}^{sub}\left(\theta,\theta_0\right) = \overline{P}_j\left(X_i;\theta_0\right) - \overline{P}_j\left(X_i;\theta\right).$$

[9] I thank one referee for raising questions on this issue.

The next two describe the relative location of the true choice probability between the bounds in (3.2), i.e.,

$$\Delta_{ij}^{sup}(\theta_0) = \Pr(d_{ij} = 1 | X_i) - \underline{P}_j(X_i;\theta_0),$$

$$\Delta_{ij}^{sub}(\theta_0) = \overline{P}_j(X_i;\theta_0) - \Pr(d_{ij} = 1 | X_i).$$

Theorem 2 characterizes the point identification condition in terms of these quantities.

Theorem 2 (Point Identification)**.** *Suppose Assumption 1 holds. Then θ_0 is point identified if and only if for any $\theta \neq \theta_0$, there exists some $j \in \mathcal{J}$ such that the set*

$$\left\{ i \in \mathbf{I} : R_{ij}^{sup}(\theta,\theta_0) > \Delta_{ij}^{sup}(\theta_0) \ \ or\ R_{ij}^{sub}(\theta,\theta_0) > \Delta_{ij}^{sub}(\theta_0) \right\} \tag{3.4}$$

has positive measure.

Proof. First, observe that θ_0 is point identified if and only if for any $\theta \neq \theta_0$, some inequalities in (3.2) are violated for a positive measure of consumers. Then the conclusion follows immediately after noting the facts,

$$R_{ij}^{sup}(\theta,\theta_0) > \Delta_{ij}^{sup}(\theta_0) \Leftrightarrow \underline{P}_j(X_i;\theta) > \Pr(d_{ij} = 1 | X_i),$$

and

$$R_{ij}^{sub}(\theta,\theta_0) > \Delta_{ij}^{sup}(\theta_0) \Leftrightarrow \overline{P}_j(X_i;\theta) < \Pr(d_{ij} = 1 | X_i).$$

□

It is instructive to consider the special case in which the bounds on choice set collapse, i.e., $C_i = C_i^{sub} = C_i^{sup}$, for a positive measure of consumers (not all the consumers). For these consumers and any $j \in C_i^{sub}$, $\Delta_{ij}^{sub}(\theta_0) = \Delta_{ij}^{sup}(\theta_0) = 0$ and the identification condition (3.4) boils down to the standard identification condition in traditional discrete choice models, i.e.,

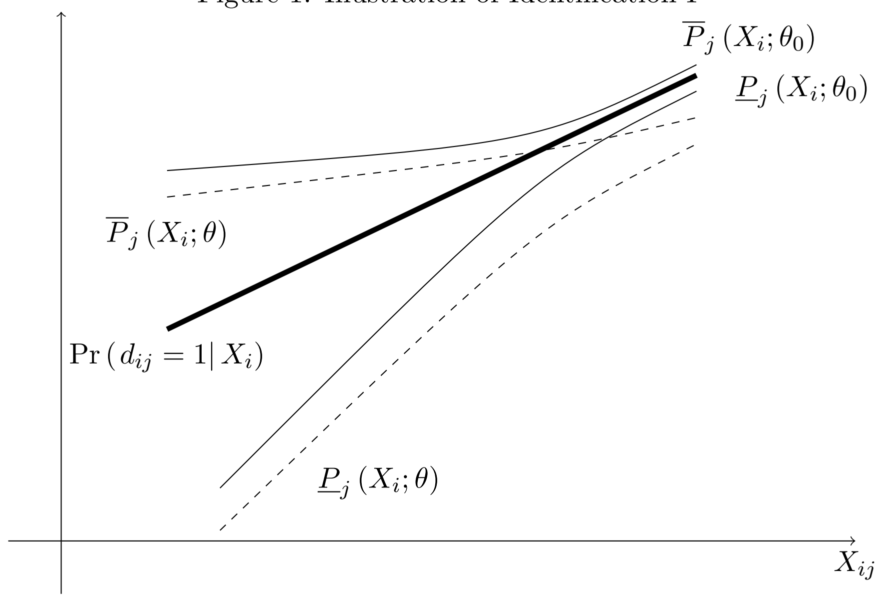
$$\left| \sigma_j\left(X_i, C_i^{sub};\theta\right) - \sigma_j\left(X_i, C_i^{sub};\theta_0\right) \right| > 0.$$

Thus, in this case, the model is point identified by the consumers with collapsed bounds, but the information from other consumers, i.e., those with $C_i^{sub} \subset C_i^{sup}$, could still be helpful for the estimation, see, e.g., Moon and Schorfheide (2006).

In general, we may not have consumers with collapsed bounds in the sample but the identification condition in Theorem 2 could still be justified in many case. In the following I shall show two such cases with graphical illustrations of the mechanisms underlying Theorem 2. For simplicity, let us focus on the products in C_i^{sub} for each consumer i, and note that $C_i^{sub} = C_i^{sub} \cup \{j\}$ for any $j \in C_i^{sub}$.

The first one, shown in Figure 1, extends the intuition in the case where a positive measure of consumers have collapsed bounds. We can see that the bounds at the true value are tight for the consumers/products with large X, thus both $\Delta_{ij}^{sup}(\theta_0)$ and $\Delta_{ij}^{sub}(\theta_0)$ are close to zero for these consumers/products. When we deviate from the true value (from θ_0 to θ), the bounds shift down; if $R_{ij}^{sub}(\theta, \theta_0) > \Delta_{ij}^{sub}(\theta_0)$, then the upper bounds for the consumers/products with large X are violated and we achieve identification. Alternatively, if the lower bound shifts up as we move from θ_0 to θ and $R_{ij}^{sup}(\theta, \theta_0) > \Delta_{ij}^{sup}(\theta_0)$, then we could also disentangle θ from θ_0.

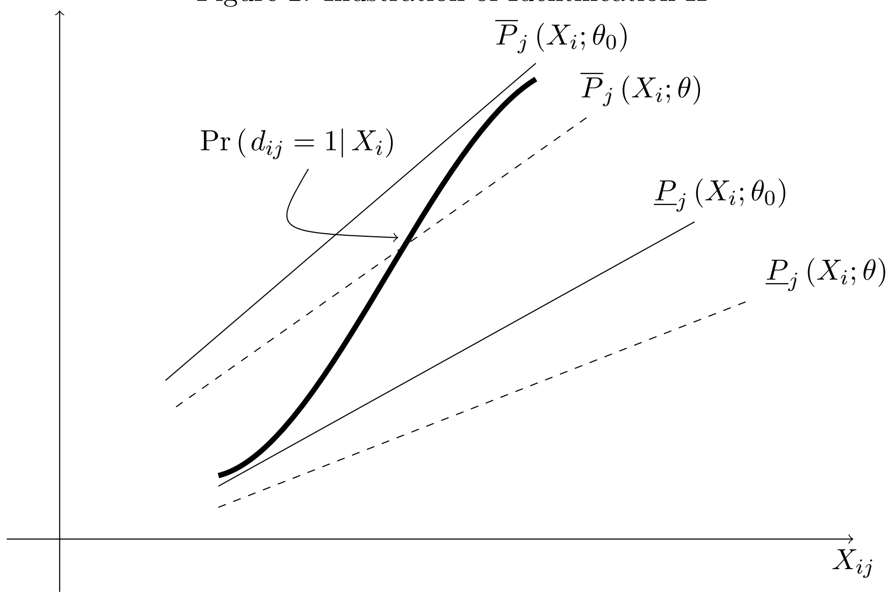
Figure 1: Illustration of Identification I



In fact, it can be quite natural that the bounds on choice probabilities for some products/consumers are much tighter than other products. For example, if for some reason consumer i likes product j much more than any other products, i.e., consumer i has a very inelastic demand for j), then i's choice probability of j is affected little by adding or dropping other products in the choice set, which means both $\Delta_{ij}^{sup}(\theta_0)$ and $\Delta_{ij}^{sub}(\theta_0)$ are small. On the other hand, the demand for some products might be very elastic so the choice probabilities are sensitive to the inclusion of other products in the choice set; in this case, the bounds on choice probabilities are wide and thus not informative.

From the above analysis, it might appear that the bounds has to be narrow, at least for some consumers/products, to achieve identification. But the next example shows that actually this is not necessary - identification could be obtained even if all the bounds are wide. The key requirement is that some consumers' true choice sets are close to one side of the bounds. In Figure 2, the true choice probabilities for the large and small X's are close to the one side of the bounds. As we can see, once we deviate from θ_0, the bounds shift and those consumers who are close to the boundary have a good chance of violating associated inequalities. And this does not rely on bounds being narrow, i.e., Δ_{ij}^{sub} or Δ_{ij}^{sup} could be small even when the bounds are wide. In the Monte Carlo section, I use this intuition to construct the data generating process so that identification could be achieved.

Figure 2: Illustration of Identification II



From the above example, we can see that we do need are informative bounds, but the bounds do not have to be narrow to be informative - they are informative as long as one side of them is close to the true choice set. Certainly, whether the identification condition could be satisfied also depends crucially on the direction and magnitude of the movement of bounds on choice probabilities (measured by $R_{ij}^{sup}(\theta,\theta_0)$ and $R_{ij}^{sub}(\theta,\theta_0)$) when the parameter departs from its true value. For example, if both bounds move away from each other as parameter shifts from θ_0 to θ, then we will not be able to distinguish θ from θ_0, because both $R_{ij}^{sup}(\theta,\theta_0)$ and $R_{ij}^{sub}(\theta,\theta_0)$ are zero.

Another important case that leads to identification failure is constructing bounds using choice-based sampling[10]. To see this, let us revisit the extreme case in which we let $C_i^{sub}=\{K_i^*\}$, where K_i^* is realized chosen product. It is clear that

$$1\left\{j\in C_i^{sub}\right\}\sigma_j\left(X_i,C_i^{sup};\theta_0\right)=\begin{cases}1, & \text{if } j=K_i^*\\ 0, & \text{otherwise}\end{cases},$$

which perfectly coincide with the observed choice variable d_{ij} for any θ. Hence, we have completely lost one side bound, so any θ such that $R_{ij}^{sub}(\theta,\theta_0)\leq 0$ can not be distinguished from θ_0. Thus, in general, we should avoid choice-based sampling in constructing C_i^{sub}. Finally, note that choice-based sampling problem is not unique to bounds approach - it leads to inconsistency in standard estimation of discrete choice models when the choice set is constructed in this way, see, e.g., Manski and Lerman (1977).

[10]I thank one referee for the helpful comments on this choice-based sampling issue.

We could also get a sense of why the standard estimator with fixed choice sets might fail from the graphical illustrations. For example, in Figure 2, if we use one side of the bounds, say upper bound $\overline{P}_j(X_i;\theta)$, as our model to estimate the parameters, then we are very likely to end up with biased estimates because the shape of the true choice probability curve is so different from $\overline{P}_j(X_i;\theta)$. Also, he Monte Carlo experiments confirm that using upper or lower bound as true choice set generates serious biases.

3.3 Estimation

The estimation is based on the inequality constraints (3.2), which can be written as a system of conditional moment inequalities

$$\begin{aligned}\mathbf{E}\left[m_j^{sub}(X_i,d_i;\theta)\middle| X_i\right] &\geq 0\\ \mathbf{E}\left[m_j^{sup}(X_i,d_i;\theta)\middle| X_i\right] &\geq 0\ \forall j\in\mathcal{J}\ a.s.\,[X_i]\,,\end{aligned}\tag{3.5}$$

where

$$m_j^{sup}(X_i,d_i;\theta)\equiv d_{ij}-1\left\{j\in C_i^{sub}\right\}\sigma_j\left(X_i,C_i^{sup};\theta_0\right),$$

and

$$m_j^{sub}(X_i,d_i;\theta)\equiv 1\left\{j\in C_i^{sup}\right\}\sigma_j\left(X_i,C_i^{sub}\cup\{j\};\theta_0\right)-d_{ij}.$$

Following Andrews and Shi (2013), the conditional moment inequalities (3.5) can be transformed into a set of unconditional moment inequalities without information loss, i.e.,

$$\begin{aligned}\mathbf{E}\left[m_j^{sub}(X_i,d_i;\theta)\,g(X_i)\right] &\geq 0\\ \mathbf{E}\left[m_j^{sup}(X_i,d_i;\theta)\,g(X_i)\right] &\geq 0\ \forall j\in\mathcal{J}\,\forall g\in\mathcal{G},\end{aligned}\tag{3.6}$$

where $\mathcal{G}$ is some space of instrumental functions [11]. Then a simple analog estimator, in the spirit of Andrews and Shi (2013), could be defined as

$$\hat{\theta}=\arg\min_{\theta\in\Theta}\int\sum_{j=1}^{J}\left\{\left(\widehat{\mathbf{E}}\left[m_j^{sub}(X_i,d_i;\theta)\,g(X_i)\right]\right)_-^2+\left(\widehat{\mathbf{E}}\left[m_j^{sup}(X_i,d_i;\theta)\,g(X_i)\right]\right)_-^2\right\}d\mu(g)\,,\tag{3.7}$$

where $\widehat{\mathbf{E}}[\cdot]\equiv\frac{1}{N}\sum_{i=1}^{N}[\cdot]$, $(x)_-=\min(0,x)$, $\mu(\cdot)$ is some measure over $\mathcal{G}$.

In the Monte Carlo experiments and the empirical application of this paper, I follow Gandhi, Lu, and Shi (2013) to define $\mathcal{G}$ as a set of indicator variables. Specifically, for any continuous variable X, define a set of indicator variables

$$\left\{g_{a,r}(X)=1\left(X\in B_{a,r}\right):B_{a,r}\in\left(\frac{a-1}{2r},\frac{a}{2r}\right],a_u\in\{1,2,...,2r\},r=1,...,\bar{r}\right\}$$

[11] See Andrews and Shi (2013) for an extensive discussion about different types of instrumental functions.

and choose $\mu(\cdot)$ such that $\mu(g_{a,r}) \propto (100+r)^{-2}(2r)^{-1}$,where $\bar{r}$ is a tuning parameter. Each categorical variable is transformed into indicator variables and assigned a common weight that is defined as $\mu \propto (100+k_c)^{-2}(2k_c)^{-1}$, where k_c is the number of categories in the variables. The final set $\mathcal{G}$ includes all the indicator variables as well as flexible interactions between them.

Note that if the point identification assumption is not desirable in some applications, this procedure could be extended to obtain a consistent estimate of the identified set by constructing a level set of the objective function [12]. For the current analysis, I focus on the point identification case. The following result shows the consistency of the estimator and the proof is in Appendix A.

Proposition 1 (Consistency)**.** *Suppose the conditions in Theorem 2 hold. Then, under regularity conditions, the estimator* (3.7) *is consistent.*

3.4 Choosing Bounds

The key inputs of the bounds approach is a pair of valid (Assumption 1) and informative (Theorem 2) bounds C_i^{sub}, C_i^{sup} on each consumer's choice set. In general, these bounds could be defined by economic theory and institutional knowledge in specific empirical applications. Here I shall provide some guidance on the choosing the bounds based on my own experiences and thoughts.

The largest set C_i^{sup} is usually not controversial and could almost always be defined as the whole market where consumer i locates, given well-defined markets in specific applications. In other words, it should be equal to the "full choice set" imposed in estimating standard discrete choice set models.

The smallest set C_i^{sub} can be tricky to pick and at the same time plays an important role for identification. First of all, according to my experiences (perhaps limited) in simulation and application, the trivial bound $C_i^{sub} = \{0\}$ usually fails to yield informative estimates, and I think it is a loss of identification, i.e., the conditions in Theorem 2 do not hold in this case [13]. To restore identification, we should use an informative lower bound C_i^{sub} that at least includes some inside good(s) besides the outside option. For example, in the Monte Carlo simulations (Section 4), each consumer has a "default" inside good j_i^* when forming his/her choice set, so I define $C_i^{sub} = \{0, j_i^*\}$ and the results show it works quite well. Also, it might be tempting to use choice-based sampling to construct C_i^{sub}, e.g., including the chosen option in C_i^{sub}, however, as we have discussed in the identification section, this might create difficulties for identification so researchers should not use this idea when constructing bounds.

Empirical researchers should always check the sensitivity of parameter estimates with respect to the choice of bounds. And if the random utility model is correctly specified, we might be able to at least perform some heuristic diagnoses on the choice of bounds. Let us again focus on the lower bound C_i^{sub}. There are three possible cases: 1) C_i^{sub} is valid but not informative enough;

[12] A formal treatment of the partial identification approach (applied to demand estimation context) could be found in Gandhi, Lu, and Shi (2013).

[13] It would be interesting to explore if the "identified set" implied by the moment inequalities in this trivial bound case is completely uninformative, i.e., partial identification. However, this is beyond the scope of this paper and is part of the future research agenda.

2) C_i^{sub} is valid and informative; 3) C_i^{sub} is not valid. In the first case, the bounds are too weak to identify the model. Thus we should expect to see a flat criterion function in (3.7) so that the parameter estimates are numerically very unstable. In the third case, the model is over-identified so the criterion function should be significantly different from zero, given other parts of the model are correctly specified. In the middle case, we should see a range of C_i^{sub} choices that generate stable parameter estimates and this is the safe zone. In one set of the Monte Carlo experiments, I explore the consequences of using too weak or strong bounds. The basic results show that it is safer to use weak bounds than to use strong bounds. And the too aggressive bounds could cause serious bias to the parameter estimates. These patterns confirm the above analysis.

Ideally, a formal statistical testing procedure and/or data-driven method should be developed. Unfortunately, the statistical/econometric theory for model selection/specification test with conditional moment inequalities has not been well developed and the current methods, e.g., Shi (2015) and Bugni, Canay, and Shi (2015), are not directly applicable to the problem in this paper. Hence, this is beyond the scope of the current paper and will be a important direction for future research.

4 Monte Carlo Simulations

This section presents simulation results for a range of Monte Carlo designs. The goal is to examine the performance of the bounds approach and compare it with several alternative estimation strategies under various scenarios.

4.1 Data Generating Process

For each consumer i, let $\mathcal{J}_i \equiv \{0, 1, ..., J_i\}$ his/her feasible set (not the final choice set), where J_i is drawn from $U\left[1, \bar{J}\right]$. Each product j in $\mathcal{J}_i$ is associated with an observed characteristic $X_{ij} \sim N\left(5 \times \frac{i}{n}, 1\right)$, where n is the total number of simulated consumers. Note that the term $\frac{i}{n}$ generates a pattern that a consumer with larger i tends to have greater X and thus stronger preference for inside goods.

The utility to consumer i from buying an inside product $j \in \mathcal{J}_i$ is

$$u_{ij} = \alpha + \beta_i X_{ij} + \epsilon_{ij},$$

where $\beta_i \sim N\left(\mu_\beta, \sigma_\beta\right)$ is a random coefficient and ϵ_{ij} is a i.i.d. Standard Gumbel random variable.

Each consumer i's constructs choice set C_i by sampling products from the feasible set $\mathcal{J}_i$. First consumer i draws the number of products $K_i \equiv |C_i|$ from one of the following distributions:

$$\text{DGP I: } K_i = \begin{cases} 1 & \frac{i}{n} \leq K_a \\ U\left[1, J_i\right] & K_a < \frac{i}{n} \leq K_b \\ J_i & \frac{i}{n} > K_b \end{cases}, \quad \text{DGP II: } K_i = \begin{cases} J_i & \frac{i}{n} \leq K_a \\ U\left[1, J_i\right] & K_a < \frac{i}{n} \leq K_b \\ 1 & \frac{i}{n} > K_b. \end{cases} \tag{4.1}$$

We could interpret K_i as a function of the underlying search cost of consumer i, i.e., low search cost generates large K, vice versa. Hence, DGP I implies that consumers with large (small) i tend to have large (small) choice sets while DGP II is in the opposite direction. Next, consumer i ranks the products in $\mathcal{J}_i$ by the value of X in descending order and includes the first K_i products in her/his choice set C_i. Finally, the outside option is included in every consumer's choice set.

We could regard this choice set generating process as a special case of fixed sample size or simultaneous search scheme (see, e.g., Moraga-González, Wildenbeest, and Sandor (2010)): first the size of search set K_i is determined and then consumer search products based on the rank of ex-ante expected utilities of products (ϵ_{ij} is not realized at the time of search).

Observe that the above data generating process creates non-trivial correlation between choice set formation and preference. A consumer with large i is likely to has large X and at the same time, she/he has systematically large (DGP I) or small (DGP II) choice set. Hence, the overall population of consumers exhibits a positive (DGP I) or negative (DGP II) correlation between the size of choice set and utility level.

Now with a well-specified utility function and choice set formation model, each consumer i's choice probabilities are obtained via simulation, i.e.,

$$\widehat{\Pr}\left(d_{ij}=1|\, X_i\right)=\frac{1}{R}\sum_{r=1}^{R}\frac{\exp\left(\alpha+\beta_r X_{ij}\right)}{1+\sum_{k\in C_i}\exp\left(\alpha+\beta_r X_{ik}\right)},$$

where $\{\beta_r : r=1,...,R\}$ is a set of random draws from $N\left(\mu_\beta,\sigma_\beta\right)$.

4.2 Estimation Strategies

To implement the bounds approach, I set the upper and lower bound as $C_i^{sup}=\mathcal{J}_i$ and $C_i^{sub}=\{0,j_i^*\}$ respectively, where $j_i^*\equiv\arg\max_{k\in\mathcal{J}_i}\{X_{ij}\}$ is the ex-ante best product of consumer i. By the design of the choice set generating process, both C_i^{sup} and C_i^{sub} are valid. And since there are a positive measure of consumers whose true choice sets are on the boundaries, i.e., equals to either C_i^{sup} or C_i^{sub}, the identification can be justified using the same argument as illustrated in Figure 2.

The oracle estimator (infeasible in practice) is obtained by estimating the standard multinomial choice model by using the true choice set C_i as observable. Also, the estimation strategies that treat the largest or smallest set as true choice sets are also included for comparisons.

To further compare the bounds approach with choice set formation models, I consider two other estimation strategies that take the choice set generation into account: independent sampling and sequential search model. The independent sampling approach follows Goeree (2008)'s algorithm, in which consumer i's choice set is simulated by the following procedure for any given value of parameter γ:

1. Draw a vector of random variables $w_i=(w_{i1},...,w_{iJ_i})$, where $w_{ij}\sim U[0,1]$;

2. For each product $j\in J_i$, include j in the choice set if and only if $\phi_{ij}>u_{ij}$, where $\phi_{ij}\equiv\frac{\exp(X_{ij}\gamma)}{1+\exp(X_{ij}\gamma)}$;

3. Collect all the products included to form consumer i's choice set C_i^{Ind}.

With the simulated choice set C_i^{Ind}, the choice probabilities are generated by the random utility model. And the parameter γ in the choice set formation is to be estimated jointly with other parameters in the model.

For the sequential search model, I use a version of Hortacsu and Syverson (2004)'s specification. Assuming equal sampling probability and the search cost distribution follows lognormal (μ_c, σ_c), the probability that consumer i chooses product j could be written as

$$P_{i1}^{seq} = \frac{1}{J_i}\left[1 - \Phi\left(\frac{\log c_{i1} - \mu_c}{\sigma_c}\right)\right]$$

$$\begin{aligned} P_{ij}^{seq} &= \frac{1}{J_i}\left[1 - \Phi\left(\frac{\log c_{i1} - \mu_c}{\sigma_c}\right)\right] + \frac{1}{J_i - 1}\left[\Phi\left(\frac{\log c_{i1} - \mu_c}{\sigma_c}\right) - \Phi\left(\frac{\log c_{i2} - \mu_c}{\sigma_c}\right)\right] \\ &\quad + ... + \frac{1}{J_i - j + 1}\left[\Phi\left(\frac{\log c_{ij-1} - \mu_c}{\sigma_c}\right) - \Phi\left(\frac{\log c_{ij} - \mu_c}{\sigma_c}\right)\right], \text{ for } j > 1, \end{aligned}$$

where $c_{ij} = \frac{1}{J_i}\sum_{k=j}^{J_i}(u_{ik} - u_{ij})$ denotes the the cutoff point on the search distribution that is derived from the optimal search model, and $\Phi(\cdot)$ is the CDF of standard normal distribution. The parameters in the search distribution, (μ_c, σ_c), are jointly estimated with the parameters in the utility function.

The above two models of choice set formation are estimated using a moment-based estimator that matches the predicted choice probabilities the observed ones. For comparison purposes, I use the same set of instrument functions $\mathcal{G}$ as the bounds approach[14].

4.3 Simulation Results

The first set of results, presented in Table 1, compare the performance of all the six estimation strategies under various scenarios of choice set heterogeneity. The labels "DGP I" and "DGP II" refers to the choice set formation model specified in (4.1). In the following, I shall focus on Panel A-I and A-II for the discussion and the results in other panels are similar. Note that in Panel A-I and A-II, I set $K_a = K_b = \frac{1}{3}$ so one third of consumers have $\mathcal{J}_i$ as their true choice set and another one third have $\{0, j_i^*\}$. Specifically, in Panel A-I, consumers with large (small) X are associated with large set $\mathcal{J}_i$ (small set $\{0, j_i^*\}$), while it is in the opposite direction in Panel A-II.

We can see that the biases of bounds approach (labeled by "Bound") are very small and close to those of the oracle estimator (labeled by "Oracle"). Also, the standard deviation of bound estimator is significantly larger than the oracle estimator, which is not surprising because the bounds approach uses less information of choice sets. All the other estimation strategies exhibit substantially larger biases and/or standard deviations than oracle and bounds estimator.

[14]See Dominguez and Lobato (2004) for a formal treatment of using indicator instrumental functions for estimating models defined by conditional moment restrictions.

An interesting pattern is that the "Full" (using $\mathcal{J}_i$ as choice set) strategy performs much worse in Panel A-II than in A-I, while the "Smallest" (using $\{0, j_i^*\}$ as choice set) shows the opposite pattern. Note that the key difference between DGP I and II is that the former (later) exhibits positive (negative) correlation between the size of choice set and utility level among the population. The intuition for why using the bounds as fixed choice sets lead to biases could be obtained by examining the graphical illustrations in Figure 1 and 2.

The choice set formation models do not perform well and I think it is simply because they are misspecified: the independence sampling approach ignores the correlation between random utility and choice set formation; the sequential search model simply can not capture the nature of fixed sample size search in the true data generating process.

Furthermore, the table also includes a comparison of the computational run-times of different estimators. We can see that the bounds approach is only slightly slower than the estimators with fixed choice sets and much faster than the choice set formation models [15].

To further explore the properties of alternative estimators, I consider a couple of extensions of the benchmark scenarios in Table 1.

1. Table 1 shows results for DGPs with different levels of choice probabilities for outside option. The choice probability of outside option is controlled by the constant term α in the utility function: a smaller α implies a larger outside choice probability, because the utility of outside option has been normalized to zero. An important pattern arises from this exercise is that the bounds approach, as well as full and smallest choice set approach, all exhibit smaller biases as the outside choice probability increases. I think the main reason is that, as the outside option becomes more important in terms of choice probability, the choice set uncertainty in the inside goods are less important so its effects on the estimates are smaller.

2. In Table 3, I examine the performance of the estimators by perturbing upper bound on the size of choice set ($\bar{J}$). Since a larger $\bar{J}$ means more uncertainty in unobserved choice sets, so we would expect larger biases and/or standard deviations of the bounds approach (the bounds become wider), as well as the estimators using fixed choice sets. The results confirm the conjecture: the biases and standard deviations of "Bound", "Full" and "Small" increase substantially as $\bar{J}$ gets larger. However, the bounds approach is still the least biased estimator (of course excluding the oracle estimator) and provides reasonably accurate estimates.

Finally, Table 4 shows the results from stress tests of bounds approach when the bounds are misspecified. In particular, I focus on the misspecification of the lower bound (smallest set) as the upper bound (largest set) seems less controversial in empirical applications. To do this, I change

[15] I thank an anonymous referee for recommending this comparison to show that the computational advantage of bounds approach over choice formation models.

the DGP slightly so that

$$\text{DGP I: } K_i = \begin{cases} \underline{J_i} & \frac{i}{n} \leq K_a \\ U\left[\underline{J_i}, J_i\right] & K_a < \frac{i}{n} \leq K_b \\ J_i & \frac{i}{n} > K_b \end{cases}, \quad \text{DGP II: } K_i = \begin{cases} J_i & \frac{i}{n} \leq K_a \\ U\left[\underline{J_i}, J_i\right] & K_a < \frac{i}{n} \leq K_b \\ \underline{J_i} & \frac{i}{n} > K_b, \end{cases} \tag{4.2}$$

where $\underline{J_i}$ is the nearest integer of $.5J_i$.

The columns labeled "BD 0" in Table 4 implement the bounds approach with lower bound correctly specified as $\underline{J_i}$. Next, the columns "BD I" and "BD II" use conservative lower bounds as $\text{Round}\,(.3J_i)$ and $\text{Round}\,(.1J_i)$, respectively. Furthermore, "BD III" and "BD IV" correspond to $\text{Round}\,(.7J_i)$ and $\text{Round}\,(.9J_i)$ as lower bounds. Hence, we can see that the lower bounds used in "BD I" and "BD II" are conservative but valid, while those used in "BD III" and "BD IV" are aggressive and misspecified. The "Oracle", "Full" and "Smallest" estimators are also included for comparison purposes.

The general lesson we could learn from the results in Table 4 is that it is much safer to have a conservative bound rather than an aggressive, misspecified lower bound. In particular, we can see that in some cases, the misspecified large lower bounds ("BD III" and "BD IV") yield even larger biases than the standard estimators with fixed choice sets. The good news is that the estimates are quite stable when conservative, weaker bounds are used ("BD I" and "BD II"). As discussed in Subsection 3.4, researchers should be cautious in choosing the bounds and always perform robustness checks to make sure the results are not sensitive.

Table 1: Monte Carlo Results I: Benchmark Cases

		DGP I						DGP II					
		Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search	Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search
		Panel A-I: $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$						Panel A-II: $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$					
$\hat{\mu}_\beta$	Bias	.0045	.0045	.1358	-1.4588	.9743	.2527	.0044	-.0025	.4892	-.3174	.4402	.3602
	S.D.	.0013	.0035	.0106	.1452	.2122	3.4300	.0018	.0037	.0413	.0198	.1011	3.3665
$\hat{\sigma}_\beta$	Bias	.0075	.0077	-.0531	.7712	-.0262	.3870	.0052	-.0316	-.2927	.1881	.2392	.3710
	S.D.	.0034	.0105	.0141	.1387	.5816	3.3204	.0039	.0192	.0516	.0154	.0897	3.3572
Ave. Runtime (s)		12.89	19.76	12.85	16.55	41.32	91.01	13.26	20.62	13.75	15.36	40.90	99.11
		Panel B-I: $K_a = 0$, $K_b = \frac{1}{2}$						Panel B-II: $K_a = 0$, $K_b = \frac{1}{2}$					
$\hat{\mu}_\beta$	Bias	.0048	-.0021	.0456	-1.6716	.2052	.3680	.0038	.0055	.5909	-.1294	.8674	.4450
	S.D.	.0022	.0037	.0056	.2206	.0347	3.3008	.0029	.0055	.0614	.0108	.2252	3.1997
$\hat{\sigma}_\beta$	Bias	.0059	-.0182	-.0242	.8556	.0149	.3868	.0036	.0065	-.3873	.0955	.1320	.3362
	S.D.	.0024	.0178	.0093	.2080	.0477	3.3363	.0058	.0110	.0644	.0126	.1162	3.2914
Ave. Runtime (s)		13.23	20.44	13.23	17.45	32.10	94.85	13.01	20.38	13.47	14.36	41.28	90.32
		Panel C-I: $K_a = \frac{1}{2}$, $K_b = 0$						Panel C-II: $K_a = \frac{1}{2}$, $K_b = 0$					
$\hat{\mu}_\beta$	Bias	.0034	-4.54E-4	.1422	-1.3329	1.1051	.4832	.0046	-.0028	.5434	-.1548	.4650	.5347
	S.D.	.0022	.0432	.0135	.1290	.2604	3.4021	.0012	.0043	.0662	.0130	.1170	3.0770
$\hat{\sigma}_\beta$	Bias	.0091	-.0054	-.0572	.7200	-.0953	.4400	.0068	-.0433	-.3296	.1259	.2366	.4253
	S.D.	.0040	.0198	.0153	.1362	.6551	3.5236	.0023	.0250	.0676	.0184	.0964	3.1342
Ave. Runtime (s)		12.69	19.34	12.86	16.13	42.09	97.18	12.80	19.20	13.24	14.02	38.88	100.66

Table 2: Monte Carlo Results II: Varying Outside Choice Probability

		DGP I with $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$						DGP II with $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$					
		Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search	Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search
		Panel I-A: $\alpha = 0$						Panel II-A: $\alpha = 0$					
$\hat{\mu}_\beta$	Bias	-4.14E-4	.0010	.1764	-1.7477	.9328	.2880	-2.12E-4	-.0268	.6187	-.3924	.5985	.2921
	S.D.	.0016	.0360	.0158	.1837	.1502	3.3498	.0011	.0412	.0781	.0237	.1005	3.4733
$\hat{\sigma}_\beta$	Bias	.0038	.0050	-.0878	.8637	.0101	.4100	.0032	-.0467	-.4080	.2372	.1351	.4384
	S.D.	.0027	.0544	.0272	.1795	.1040	3.3195	.0015	.0668	.0920	.0248	.0843	3.5282
Ave. $\Pr(d_0 = 1)$					.1299						.1080		
		Panel I-B: $\alpha = -3$						Panel II-B: $\alpha = -3$					
$\hat{\mu}_\beta$	Bias	.0049	.0012	.0565	-.7422	.7097	.3336	.0051	.0056	.2477	-.1650	.3284	.3742
	S.D.	7.05E-4	.0036	.0029	.0201	.0787	3.3784	4.40E-4	.0089	.0115	.0071	.0623	3.3946
$\hat{\sigma}_\beta$	Bias	.0126	.0126	-.0105	.1382	.2645	.3612	.0127	-.0258	-.1550	.0347	.3918	.3008
	S.D.	.0036	.0208	.0068	.0069	.0928	3.3630	.0030	.0320	.0125	.0107	.1024	3.4680
Ave. $\Pr(d_0 = 1)$					.3692						.3311		
		Panel I-C: $\alpha = -5$						Panel II-C: $\alpha = -5$					
$\hat{\mu}_\beta$	Bias	6.45E-4	-9.15E-4	.0164	-.4913	.4163	.2905	1.36E-4	.0152	.0826	-.0929	.2512	.3439
	S.D.	.0022	.0050	.0017	.0177	.0409	3.4299	8.59E-4	.0121	.0071	.0066	.0985	3.3918
$\hat{\sigma}_\beta$	Bias	.0183	.0138	.0184	.0919	.2753	.3102	.0213	-.0090	-8.59E-4	.0524	.5175	.2012
	S.D.	.0079	.0179	6.57E-4	.0096	.0811	3.5304	.0022	.0203	.0139	.0039	.1301	3.4971
Ave. $\Pr(d_0 = 1)$					.5469						.5401		

Table 3: Monte Carlo Results III: Varying Size of Choice Set

		DGP I with $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$						DGP II with $K_a = \frac{1}{3}$, $K_b = \frac{2}{3}$					
		Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search	Oracle	Bound	Full	Smallest	Ind. Search	Seq. Search
		Panel I-A: $\bar{J} = 10$						Panel II-A: $\bar{J} = 10$					
$\hat{\mu}_\beta$	Bias	-8.69E-4	.0051	.0765	-1.0367	.6360	.5147	-7.42E-4	-.0152	.2617	-.2130	.5974	.3964
	S.D.	.0019	.0234	.0076	.1041	.1012	2.7020	.0013	.0328	.0241	.0198	.1109	2.6368
$\hat{\sigma}_\beta$	Bias	.0044	.0131	-.1040	.6104	.2232	.3299	.0039	-.0299	-.4576	.1534	.2539	.2832
	S.D.	.0033	.0361	.0143	.0805	.1061	2.8558	.0017	.0549	.0500	.0126	.1086	2.8124
Ave. Runtime (s)		5.70	10.77	6.63	7.17	15.44	70.71	5.11	10.48	6.23	6.05	13.59	63.15
		Panel I-B: $\bar{J} = 50$						Panel II-B: $\bar{J} = 50$					
$\hat{\mu}_\beta$	Bias	-3.91E-4	-.0084	.2045	-2.0133	1.0743	.1822	-5.85E-5	-.0307	.6991	-.4750	.5648	.1179
	S.D.	.0030	.0454	.0356	.2439	.2010	3.5541	.0015	.0463	.1548	.0333	.0973	3.4888
$\hat{\sigma}_\beta$	Bias	.0034	-.0107	-.1004	.8865	-.0760	.3874	.0030	-.0514	-.4354	.2735	.0951	.4466
	S.D.	.0055	.0759	.0449	.2262	.1253	3.5800	.0020	.0739	.1530	.0404	.0951	3.3685
Ave. Runtime (s)		15.99	35.29	18.18	23.72	40.70	104.23	16.02	34.81	19.97	19.65	40.26	102.22
		Panel I-C: $\bar{J} = 100$						Panel II-C: $\bar{J} = 100$					
$\hat{\mu}_\beta$	Bias	-4.23E-4	-.0243	.2319	-2.2916	1.1618	.5139	1.38E-4	-.0356	.7198	-.5825	.5393	.3958
	S.D.	.0035	.0415	.0633	.2699	.2038	3.7146	.0014	.0452	.2800	.1290	.1564	3.6063
$\hat{\sigma}_\beta$	Bias	.0038	-.0371	-.0992	.8922	-.1420	.2715	.0028	-.0572	-.4065	.3422	.0386	.3163
	S.D.	.0076	.0720	.0863	.2345	.1042	3.7898	.0031	.0823	.2853	.0935	.0996	3.7052
Ave. Runtime (s)		33.56	68.95	41.24	52.18	85.33	174.66	34.98	74.95	46.46	44.90	88.47	175.37

Table 4: Monte Carlo Results IV: Misspecified Bounds

		DGP I with $K_a = \frac{1}{3}, K_b = \frac{2}{3}$								DGP II with $K_a = \frac{1}{3}, K_b = \frac{2}{3}$							
		Oracle	Full	Smallest	BD 0	BD I	BD II	BD III	BD IV	Oracle	Full	Smallest	BD 0	BD I	BD II	BD III	BD IV
		Panel I-A: $\bar{J} = 20$, $\alpha = 0$								Panel II-A: $\bar{J} = 20$, $\alpha = 0$							
$\hat{\mu}_\beta$	Bias	-5.32E-4	.0808	-1.5391	-8.21E-4	.0048	.0045	.1070	.0675	-6.60E-4	.1933	-.3933	-.0202	-.0210	-.0048	-.0448	-.0309
	S.D.	.0015	.0124	.1879	.0053	.0214	.0062	.0453	.0297	.0013	.0179	.0330	.0316	.0352	.0050	.0710	.0960
$\hat{\sigma}_\beta$	Bias	.0038	-.0910	.8084	.0080	.0144	-.0033	.1140	-.9779	.0034	-.2299	.2379	-.0367	-.0380	-.0169	-.5148	-.6868
	S.D.	.0027	.0187	.1596	.0075	.0338	.0161	.4778	.1658	.0021	.0409	.0292	.0495	.0554	.0211	.3712	.5326
		Panel I-B: $\bar{J} = 50$, $\alpha = 0$								Panel II-B: $\bar{J} = 50$, $\alpha = 0$							
$\hat{\mu}_\beta$	Bias	1.30E-4	.1207	-2.1020	-3.80E-4	.0038	.0038	.0890	.1772	1.30E-4	.2603	-.6487	-.0305	-.0313	-.0078	-.1587	-.0878
	S.D.	.0038	.0194	.2439	.0081	.0201	.0074	.0496	.0200	.0022	.0473	.0929	.0456	.0485	.0065	.0708	.0937
$\hat{\sigma}_\beta$	Bias	.0035	-.0439	.9096	.0089	.0122	-.0044	.0324	-.9977	.0033	-.1379	.3551	-.0504	-.0512	-.0222	-.6166	-.4707
	S.D.	.0044	.0238	.2313	.0121	.0312	.0108	.4665	.0408	.0042	.0590	.1093	.0718	.0749	.0155	.4312	.6096
		Panel I-C :$\bar{J} = 20$, $\alpha = -5$								Panel II-C: $\bar{J} = 20$, $\alpha = -5$							
$\hat{\mu}_\beta$	Bias	2.14E-4	.0109	-.4235	-.0017	-.0015	3.54E-4	-.0191	.0365	-1.67E-4	.0430	-.0716	.0136	.0126	.0096	.0469	.0674
	S.D.	.0011	.0014	.0248	.0014	.0018	.0024	.0230	.0173	.0011	.0023	.0092	.0086	.0087	.0069	.0127	.0119
$\hat{\sigma}_\beta$	Bias	.0241	.0192	.0985	.0147	.0152	.0148	.4723	.3057	.0219	.0063	.0548	-.0031	-.0014	-.0056	-.1070	-.2039
	S.D.	.0048	.0022	.0134	.0111	.0150	.0108	.0844	.0828	.0029	.0053	.0054	.0089	.0123	.0115	.0629	.0635
		Panel I-D :$\bar{J} = 50$, $\alpha = -5$								Panel II-D: $\bar{J} = 50$, $\alpha = -5$							
$\hat{\mu}_\beta$	Bias	.0035	.0202	-.5988	-.0035	-.0036	-6.87E-4	.0329	.1530	.0032	.0523	-.1391	.0180	.0174	.0158	-7.15E-4	.0182
	S.D.	.0012	.0022	.0237	.0101	.0155	.0112	.0487	.0197	9.25E-4	.0026	.0101	.0152	.0156	.0131	.0191	.0173
$\hat{\sigma}_\beta$	Bias	.0212	.0173	.1116	.0144	.0149	.0068	.1240	-.3179	.0225	-.0060	.0504	-.0235	-.0248	-.0202	-.0757	-.2077
	S.D.	.0046	.0016	.0106	.0379	.0423	.0486	.1659	.1233	.0046	.0080	.0091	.0289	.0324	.0284	.0901	.0744

5 Empirical Application

5.1 Data Description

I obtain the household purchase panel and store data from Information Resources Inc. (IRI).[16] The household panel data keeps track of the purchase history of products in 30 categories for a sample of households in two cities, Eau Claire, WI and Pittsfield, MA over the years 2001-2011. For each household in the sample, the data records the timing, location and products bought on every shopping trips to the set of stores that IRI data set covers (about 80% of all the stores in the two cities). The store data contain information on quantity, price and marketing mix at store/week/UPC level. Combining household panel with store data, we could obtain the price, marketing mix for the set of UPCs that are available on shelf for the store/week at which each purchase occurred in the household panel data.

For the current empirical exercise, I focus on the potato chips category, however, the method proposed in this paper is applicable to many other industries and by no means restricted to the category selection here. Potato chips come with different brands, flavors and package sizes; usually there is a large varieties of potato chips UPCs sold in a given store/week. A "product" is defined as a brand-size combination because the prices and promotion activities of UPCs under the same combination often move together, see also Hendel and Nevo (2006), etc.

The UPC level price, display and feature advertising are aggregated into product level for each store/week. Specifically, the price of a product is a sale-weighted price index constructed from UPC level prices. Also, display (feature) is an indicator variable that equals to 1 if any UPC within the product is on display (has feature advertising).

The sample consists of household shopping trips in Eau Claire and Pittsfield during the last 6 months of 2011. The inside goods are defined as the top 24 products that makes up more than 90% of all the purchases of potato chips in the sample. Outside option includes other products outside the top 24 and non-purchase of any potato chips. Also, the households that bought less than 5 packs of potato chips during January 2009 to June 2011 (30 months) are dropped from the sample as they are unlikely in the market of potato chips. This leaves us 2454 households with $34,150$ shopping trips in the sample.

For each shopping trip in the sample, I observe the household identifier, week, store, product purchased. Combing with household purchase history and store data corresponding to the sample period, we could obtain the following information for each purchase occasion:

1. Price, display and feature for the set of products available at the week/store of the purchase;
2. Product purchased;
3. Purchase history of the household that makes the current purchase.

Table 5 shows the summary statistics for some products in the sample. We can see that Lays is a dominant brand in the market - it is sold more than other brands in total. The price in general

[16] See Bronnenberg, Kruger, and Mela (2008) for an overview of the IRI data set.

is lower for larger size packs, showing the stylized pattern of price discrimination by size. Also, since the price includes the effect of price cut promotion, which is often associated with feature advertising, so higher price usually corresponds to fewer feature. In addition, display and feature exhibits interesting variations, e.g., Lays Natural-8.5oz is much displayed but has little feature advertising.

Table 5: Summary Statistics for Selected Products

Product (Brand-Size)	No. of Purchases	Ave. Price (Per 16oz)	Ave. Display	Ave. Feature
LAYS-10oz	907	5.50	0.34	0.36
LAYS NATURAL-10.5oz	782	5.14	0.48	0.38
WAVY LAYS-10.5oz	646	5.38	0.49	0.19
LAYS NATURAL-8.5oz	309	6.11	0.42	0.08
BARREL O FUN-12.25oz	229	3.34	0.61	0.15
BARREL O FUN-14oz	223	2.93	0.60	0.17
OLD DUTCH-10oz	177	5.36	0.08	0.11
BARREL O FUN-10oz	123	4.18	0.54	0.16
CAPE COD-8oz	112	6.73	0.07	0.11
RUFFLES-9.5oz	97	6.68	0.17	0.14

5.2 Empirical Strategies and Results

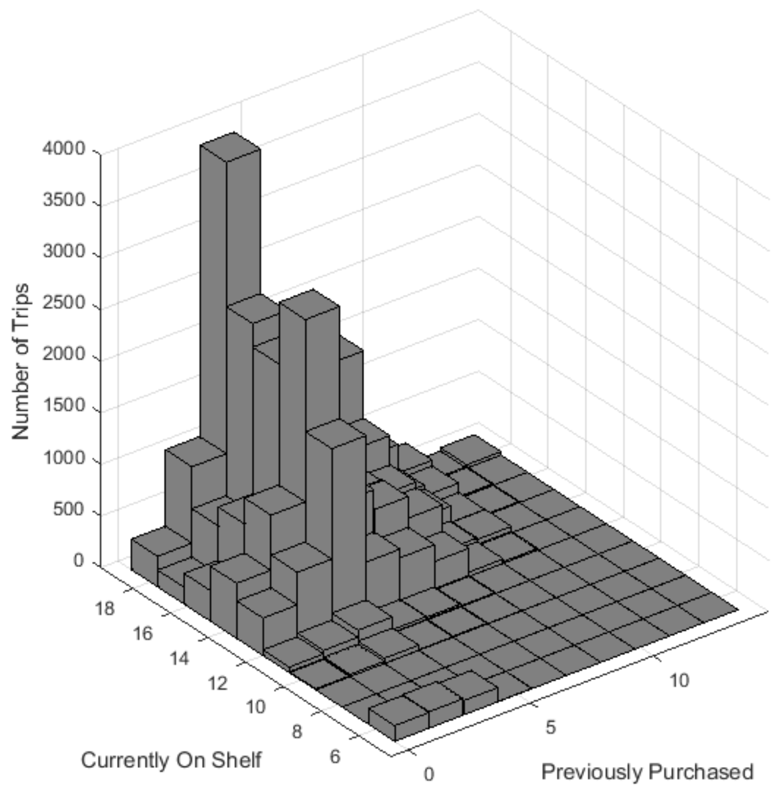
The actual choice set of a consumer in a particular shopping trip is not observed in the data. But a reasonable upper bound (largest choice set) on the choice set could be defined as all the products available on the shelf at the store/week that the purchase happens. This rules out the possibility that consumers substitute to another store/week if certain brands are not available at the current store/week. Since the potato chips are the so-called frequently-purchased-products and not expensive, this assumption seems sensible in this setting and is likely to hold for most consumers in the sample.

The lower bound (smallest choice set) is assumed to be the products available on the shelf (i.e., within the upper bound), and that a consumer ever bought during January 2009 to June 2011, i.e., 30 months prior to the sample period. The rationale for this assumption is that consumers should have learned their preferences for the potato chips they have consumed in the recent couple of years. The choice of 30 months looks arbitrary but alternative lengths of period, say 12, 24, 36, yield almost identical lower bounds for most of the households in the sample. This is not surprising because most consumers in the sample are very loyal to a small number (mostly one or two) of brand-size combinations, which is also a silent feature of the scanner data in general, see Keane (2013).

Figure 3 depicts the pattern of the bounds in the data. The axis labeled by "Previously Purchased" indicates the size of the lower bound, while the one labeled by "Currently On Shelf" is for the size of the upper bound. We can see that most of the shopping trips in the sample happens in the area where the lower bound has 0-8 products and the upper bound has 13-20 products.

Also, the figure also indicates decent variations in both bounds, although only the sizes, not the compositions, of them are shown.

Figure 3: Number of Trips for Different Sizes of Choice Set



Let i and t index consumer and purchase occasion, respectively. I estimate a standard random coefficient logit model, in which the utility to consumer i of buying j at t is modeled as

$$u_{ijt} = \beta_i X_{jt} - \alpha_i p_{jt} + \delta_j + \epsilon_{ijt},$$

where δ_j denotes the product level effect for j, p_{jt} is the sale-weighted price index for product j at t, X_{jt} includes indicators for in-store display and feature advertising, $\left(\alpha_i, \beta_i'\right)$ is a vector of mutually independent normally distributed random coefficients and ϵ_{ijt} is a i.i.d. Standard Gumbel random variable. The utility of outside option is normalized to $u_{i0t} = \delta_0 + \epsilon_{i0t}$, where δ_0 is a constant to be estimated (mainly identified by the variation in the overall level of choice probabilities across market).

The upper and lower bounds on each consumer's choice set are defined above and other implementation details are the same as Section 4 (the Monte Carlo section). The point estimates and standard errors (obtained using a standard subsampling procedure as suggested by e.g., Andrews

(1999)) are presented under the column labeled as "Bound" in Table 6. For comparison, the table also includes results (point estimates and standard errors obtained using standard MLE) from two alternative specifications with fixed choice sets. The first column labeled by "Full" presents the results from the specification assuming that each consumer has all the products available on the shelf in his/her choice set for each purchase occasion, i.e., the upper bound adopted by the bounds approach. The second column, labeled by "Smallest", is for the case in which the choice set of each consumer in each purchase occasion only includes the ones that were purchased during the 30 months and are currently available on the shelf, as well as the chosen option. Finally, the results of the three estimator for simple logit model are also included.

Table 6: Estimation Results

		Random Coefficient Logit			Simple Logit		
		Full	Smallest	Bound	Full	Smallest	Bound
Mean Coefficient							
Price		−.76 (.0009)	−.58 (.0009)	−.77 (.02)	−.50 (.0001)	−.43 (.0001)	−.28 (.02)
Display		.002 (.0002)	.003 (.0002)	.001 (.03)	.16 (.0002)	.14 (.0002)	.16 (.04)
Feature		.38 (.0002)	.34 (.0002)	.33 (.04)	.20 (.0002)	.28 (.0002)	.45 (.04)
Pre. Purch.		1.99 (.0002)	-	-	2.12 (.0002)	-	-
Product Dummy							
LAYS-10oz		.18 (.0006)	.24 (.0006)	3.90 (.10)	.44 (.0007)	.40 (.0006)	.22 (.11)
LAYS NATURAL-10.5oz		.0013 (.0006)	.0012 (.0005)	3.72 (.10)	.19 (.0006)	.25 (.0006)	.22 (.10)
WAVY LAYS-10.5oz		−.08 (.0006)	.07 (.0005)	3.58 (.10)	.12 (.0006)	.08 (.0005)	.10 (.11)
LAYS NATURAL-8.5oz		−.06 (.0007)	−.09 (.0006)	3.23 (.12)	.27 (.0008)	.25 (.0007)	−.04 (.13)
BARREL O FUN-12.25oz		−1.06 (.0005)	−.88 (.0005)	2.46 (.10)	−1.22 (.0005)	−1.08 (.0005)	−.59 (.11)
Outside Option (δ_0)		2.39 (.0008)	1.24 (.0009)	4.64 (.01)	3.15 (.0005)	1.37 (.0005)	2.25 (.02)
Other Products		Included (Estimates Omitted)					
Std. Dev. of Coefficient							
Price		.25 (.0002)	.19 (.0002)	.36 (.03)			
Display		.16 (.0001)	.12 (.0001)	.87 (.02)	-	-	-
Feature		.15 (.0001)	.15 (.0001)	.31 (.02)			
Price Elast. (%)	Mean	-4.28	-3.27	{-4.35, -4.31}	-2.89	-2.44	{-1.58, -1.55}
	Std. Dev.	1.75	1.33	{2.31, 2.33}		-	
Display Semi-Elast. (%)	Mean	.43	.47	{1.45, 1.96}	15.93	14.16	{15.73, 15.39}
	Std. Dev.	15.80	11.86	{86.28, 83.98}		-	
Feature Semi-Elast. (%)	Mean	37.11	33.14	{31.61, 30.76}	17.80	27.31	{44.73, 43.78}
	Std. Dev.	14.75	14.75	{30.54, 29.88}		-	

The mean coefficients for marketing mix variables (price, display and feature advertising) are remarkably close among the three estimators for the random coefficients model (except that the "Smallest" estimator yields a slightly smaller price coefficient). The major differences among the three approaches are in the estimated standard deviations of the random coefficients on marketing mix variables.[17] Specifically, comparing to the other two estimators with fixed choice sets, the bounds approach implies significantly larger consumer heterogeneity in the sensitivity to these marketing mix variables. Furthermore, the differences in the consumer heterogeneity could be seen from the standard deviations of price and advertising elasticities (due to the random coefficient specification) in the bottom panel of Table 6 (bounds approach has two elasticities in $\{\cdot\}$ for the two bounds on choice probabilities respectively). Finally, note that overall the results from random coefficient logit model are rather different from simple logit specification, which might suggest potential misspecification of simple logit model.

Lastly, Table 7 summarizes the Mean Squared Error (MSE) and Mean Absolute Error (MAE) of predicted choice probabilities (against actual choices) for a set of moments, from which we could get a sense of how the models fit the data. The MSE and MAE for bounds approach measures the violation of bounds and are defined as $\frac{1}{NJ}\sum_{i,j}\left\{\left[m_j^{sub}\left(X_i, d_i; \theta\right)\right]_-^2 + \left[m_j^{sup}\left(X_i, d_i; \theta\right)\right]_-^2\right\}$ and $\frac{1}{NJ}\sum_{i,j}\left\{\left|\left[m_j^{sub}\left(X_i, d_i; \theta\right)\right]_-\right| + \left|\left[m_j^{sup}\left(X_i, d_i; \theta\right)\right]_-\right|\right\}$. Though the MSE and MAE for bounds may not be directly comparable to those for standard estimators, we could at least see that the bound estimator fits the data pretty well in terms of the moments shown in the table. Also, it is interesting to note that the prediction error of simple logit is even smaller than the random coefficient logit model, however, notice that this is just a limited set of moments in the model - the more general random coefficient model should have matched many other features in the data.

[17] Another disparity is that the bounds approach gives substantially higher estimates for the product-level effects in absolute levels; however, since only differences in the product-specific effects are relevant for interpreting the model, the actual discrepancy is not large.

Table 7: Goodness of Fit: Measuring Prediction Errors

	Group			Random Coefficient Logit			Simple Logit		
Pre. Pur.	Feature	Display		Full	Smallest	Bound	Full	Smallest	Bound
Yes	Yes	Yes	MSE	.0557	.0555	.0001	.0545	.0545	1.33E-5
			MAE	.1052	.1123	.0010	.1119	.1113	.0008
Yes	Yes	No	MSE	.0458	.0458	8.56E-5	.0444	.0444	8.98E-6
			MAE	.0853	.0936	.0008	.0860	.0868	.0006
Yes	No	Yes	MSE	.0211	.0211	3.57E-5	.0207	.0207	9.45E-7
			MAE	.0363	.0417	.0003	.0405	.0409	.0001
Yes	No	No	MSE	.0138	.0138	1.72E-5	.0134	.0134	5.85E-7
			MAE	.0254	.0304	.0002	.0267	.0277	7.58E-5
No	Yes	Yes	MSE	.0076	.0077	.0001	.0076	.0077	3.79E-5
			MAE	.0149	.0077	.0006	.0142	.0077	.0005
No	Yes	No	MSE	.0056	.0056	5.97E-5	.0055	.0056	2.16E-5
			MAE	.0117	.0056	.0004	.0103	.0056	.0003
No	No	Yes	MSE	.0024	.0024	1.78E-5	.0024	.0024	3.26E-6
			MAE	.0046	.0024	9.30E-5	.0046	.0024	.7.92E-5
No	No	No	MSE	.0012	.0012	6.79E-6	.0012	.0012	1.53E-6
			MAE	.0030	.0012	4.64E-5	.0026	.0012	3.66E-5
	Overall		MSE	.0078	.0079	2.52E-5	.0077	.0077	5.05E-6
			MAE	.0149	.0146	.0002	.0153	.0140	.0001

6 Concluding Remarks

In this paper, I present a bounds approach to estimating random utility model when consumers' choice sets are unobserved and heterogeneous. Comparing to the various choice set generation models in the literature, the bounds approach uses intuitive restrictions on the choice set distribution without modeling choice set formation and is very easy to implement. Monte Carlo and empirical results show that the bounds approach is useful in correcting the biases caused by misspecified choice sets/choice set formation in the model.

Finally, I have developed the bounds approach in the context of individual choice data. But the choice set heterogeneity could also be a problem for aggregate data. In an on-going project, I consider the extensions of the current framework to handle aggregate data, in which case price endogeneity becomes an important concern. Also, future research that applies the proposed approach to empirical applications to the markets with many choices, like books, hotels, air tickets, etc., may be fruitful. Since consumers in these markets usually do not consider all the available options, the bounds approach is useful in understanding and estimating the demand in these markets.

A Proof of Consistency

In the following, I present a consistency theorem (more general than Proposition 1) that allows for unknown functions in the generic moment inequalities model

$$\mathbf{E}\left[m\left(W_i;\theta,\tau\right)\mid X_i\right]\geq 0\;\forall\tau\in\mathcal{T}\quad a.s.\left[X_i\right], \tag{A.1}$$

where $\theta\left(\equiv(\lambda,h)\right)\in\Theta\left(\equiv\Lambda\times\prod_{j=1}^{J}\mathcal{H}_j\right)$ and $\mathcal{T}$ is an index set that can be finite or infinite. This general setting is useful when empirical researchers would like to include non-parametric components in the model. Suppose θ includes nonparametric components $h(\cdot)\equiv\left(h_1(\cdot),...,h_J(\cdot)\right)$ and I shall use sieve method to approximate the unknown functions h. To keep things simple, I restrict the attention to the smooth class of functions as the parameter space for h. Specifically, define a Hölder ball with smoothness α as

$$\mathbf{H}_M^{\alpha}\left(\mathcal{X}\right)=\left\{f\in C^k\left(\mathcal{X}\right):\sup_{k\leq\underline{\alpha}}\sup_{X\in\mathcal{X}}\left|D^kf\left(X\right)\right|\leq M,\sup_{k=\underline{\alpha}}\sup_{X\neq X'}\frac{\left|D^kf\left(X\right)-D^kf\left(X'\right)\right|}{\left\|X-X'\right\|_E^{\alpha-\underline{\alpha}}}\leq M\right\},$$

where $\underline{\alpha}$ is the greatest integer smaller than α, $\|\cdot\|_E$ is the Euclidean norm, $q\in(0,1]$, $\mathcal{X}$ denotes the support of X_i, $C^k\left(\mathcal{X}\right)$ is the space of all m-times continuously differentiable real-valued functions on $\mathcal{X}$, $k\equiv k_1+k_2+...+k_{d_X}$ and $D^{\alpha}\equiv\frac{\partial^{[\alpha]}}{\partial X_1^{\alpha_1}...\partial X_{d_X}^{\alpha_{d_X}}}$ denotes the differential operator. Define a metric on Θ, $d\left(\theta,\theta'\right)=\left\|\lambda-\lambda'\right\|_E+\sum_{j=1}^{J}\left\|h_j-h_j'\right\|_H$, where $\|\cdot\|_{\mathcal{H}}$ denotes the usual sup-norm or L_2-norm.

A.1 Assumptions

Assumption 2. *The data* $\{W_i\}_{i=1}^n$ *is i.i.d.*

Assumption 3. *The parameter space of the finite dimensional parameters is compact and* $\mathcal{H}_j=\mathbf{H}_M^{\alpha}\left(\mathcal{X}\right)$ *for all* $j=1,...,J$.

Assumption 4. $\mathbf{E}\left[\sup_{\theta\in\Theta}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m\left(W_i;\theta,\tau,g\right)\right|\right]<\infty$ *and* $\mathbf{Var}\left[\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m\left(W_i;\theta_0,\tau,g\right)\right|\right]<\infty$.

Assumption 5. *There exists a constant* $c>0$ *and a random variable* $U_1\left(W_i\right)$ *satisfying* $\mathbf{E}\left[U_1\left(W_i\right)\right]^2<\infty$ *such that for any* θ,$\theta'\in\Theta$,

$$\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m\left(W_i;\theta,\tau,g\right)-m\left(W_i;\theta',\tau,g\right)\right|\leq U_1\left(W_i\right)\cdot\left[d\left(\theta,\theta'\right)\right]^c.$$

Assumption 6. *For any* $\theta\in\Theta$, *there exists* $\Pi_n\theta\in\Theta_n$ *such that* $d\left(\Pi_n\theta,\theta\right)=o(1)$;

Remark 3. Assumption 2 rules out interactions among consumers but is typically imposed in the discrete choice literature. Assumption 3 implies that Θ is compact under $d(\cdot,\cdot)$, which is a common requirement in the non-/semi-parametric literature, see, e.g., Newey and Powell (2003); Ai and Chen (2003). Allowing for non-compact parameter space is technically more involved (see, e.g., Chen and Pouzo (2012)) and seems unnecessary for the current application. Assumption 4 and 5 are some basic requirements on moment functions and are typically imposed even in the nonlinear parametric estimation literature. Assumption 6 simply defines the sieve space, see, e.g., Newey and Powell (2003); Ai and Chen (2003); Chen (2007) for more explanations.

Theorem 3 (Consistency)**.** *Suppose the conditions in Theorem 2 hold. If Assumption 2, 3, 4, 5, 6 hold, then* $d\left(\hat{\theta},\theta_0\right)=o_p(1)$.

A.2 Proof of Theorem 3

Let $Q(\theta)=S[m(\theta)]$, where $m(\theta)=\{\mathbf{E}[m(W_i;\theta,\tau,g)]:(\tau,g)\in\mathcal{T}\times\mathcal{G}\}$. Let $Q_n(\theta)=S[m_n(\theta)]$ and thus the estimator can be re-written as $\hat{\theta}\equiv\arg\min_{\theta\in\Theta_n}Q_n(\theta)$. I first present a proposition that will be used to show consistency. This proposition combines and rephrases the Remark 3.1 (4) in Chen (2007) and Lemma A1 of Newey and Powell (2003).

Proposition 2. *Suppose the following conditions hold: (i) Θ is compact under $d(\cdot,\cdot)$; (ii) $Q(\theta)$ is lower semicontinuous on Θ under $d(\cdot,\cdot)$; (iii) $Q(\theta)$ has a unique minimum on Θ at $\theta_0\in\Theta$ and $Q(\theta_0)<\infty$; (iv) the sieve spaces $\{\Theta_n:n\geq 1\}$ are compact subsets of Θ under $d(\cdot,\cdot)$ and for each $\theta\in\Theta$, there exists $\Pi_n\theta\in\Theta_n$ such that $d(\Pi_n\theta,\theta)=o(1)$; (v) $\sup_{\theta\in\Theta}|Q_n(\theta)-Q(\theta)|=o_p(1)$. Then $d\left(\hat{\theta},\theta_0\right)=o_p(1)$.*

The following lemma shows (v) of Proposition 2, i.e., uniform convergence of $Q_n(\cdot)$.

Lemma 1. *If Assumptions 2, 3, 4, 5 hold, then* $\sup_{\theta\in\Theta}|Q_n(\theta)-Q_F(\theta)|=o_p(1)$.

Proof. For any $\delta>0$, since Θ is compact by Assumption 3, we can construct a finite cover of Θ, $B\left(\theta^i,\delta\right)=\left\{\theta\in\Theta:d\left(\theta,\theta^i\right)\leq\delta\right\}$ $i=1,...,N_\delta$. For each $B\left(\theta^i,\delta\right)$, we have

$$\begin{aligned}
&\sup_{\theta\in B(\theta^i,\delta)}|Q_n(\theta)-Q_F(\theta)|\\
\leq&\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|[m_n(\theta,\tau,g)]_-^2-[m_F(\theta,\tau,g)]_-^2\right|\\
\leq&\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left\{|m_n(\theta,\tau,g)-m_F(\theta,\tau,g)|^2+2|m_F(\theta,\tau,g)||m_n(\theta,\tau,g)-m_F(\theta,\tau,g)|\right\}\\
\leq&\left[\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}|m_n(\theta,\tau,g)-m_F(\theta,\tau,g)|\right]^2\\
&+2\left[\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}|m_F(\theta,\tau,g)|\right]\left[\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}|m_n(\theta,\tau,g)-m_F(\theta,\tau,g)|\right].
\end{aligned}$$

Next, note that

$$
\begin{aligned}
&\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n(\theta,\tau,g)-m_F(\theta,\tau,g)\right|\\
\leq&\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n(\theta,\tau,g)-m_n\left(\theta^i,\tau,g\right)\right|+\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n\left(\theta^i,\tau,g\right)-m_F\left(\theta^i,\tau,g\right)\right|\\
+&\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F\left(\theta^i,\tau,g\right)-m_F(\theta,\tau,g)\right|\\
\leq&\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m(W_i;\theta,\tau,g)-m\left(W_i;\theta^i,\tau,g\right)\right|\\
+&\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n\left(\theta^i,\tau,g\right)-m_F\left(\theta^i,\tau,g\right)\right|+C_\delta\cdot\delta^c\\
\leq&\frac{1}{n}\sum_{i=1}^{n}U_1(W_i)\cdot\delta^c+\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n\left(\theta^i,\tau,g\right)-m_F\left(\theta^i,\tau,g\right)\right|+C_\delta\cdot\delta^c\\
\leq&\delta^c\left|\frac{1}{n}\sum_{i=1}^{n}U(W_i)-\mathbf{E}_F\left[U(W_i)\right]\right|+C^{'}+\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_n\left(\theta^i,\tau,g\right)-m_F\left(\theta^i,\tau,g\right)\right|+C_\delta\cdot\delta^c\\
=&o_p(1)+C_\delta\cdot\delta^c+C^{'}
\end{aligned}
$$

for some finite $C^{'}, C_\delta>0$, where the second and third inequalities follow by Assumption 5, the last equality follows from the i.i.d. assumption and Assumption 4. Also, by Assumption 4,

$$
\sup_{\theta\in B(\theta^i,\delta)}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F(\theta,\tau,g)\right|\leq\mathbf{E}_F\left[\sup_{\theta\in\Theta}\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m(W_i;\theta,\tau,g)\right|\right]<M_\delta
$$

for some finite $M_\delta>0$. Hence,

$$
\begin{aligned}
\sup_{\theta\in\Theta}\left|Q_n(\theta)-Q_F(\theta)\right| &\leq \max_i\sup_{\theta\in B(\theta^i,\delta)}\left|Q_n(\theta)-Q_F(\theta)\right|\\
&= \left(C_\delta\delta^c+C^{'}\right)^2+2M_\delta\left(C_\delta\delta^c+C^{'}\right)+o_p(1)\\
&\equiv R(\delta)+o_p(1).
\end{aligned}
$$

Finally, for any $\epsilon>0$, there exists $\delta\equiv R^{-1}(\epsilon)>0$ such that

$$
\sup_{\theta\in\Theta}\left|Q_n(\theta)-Q_F(\theta)\right|\leq\epsilon+o_p(1).
$$

The conclusion follows by noting that ϵ is arbitrary. □

Proof of Theorem 3

The proof amounts to verifying the conditions of Lemma 2. Condition (i) is implied by Assumption 3. To show condition (ii), observe that

$$\begin{aligned}
&\left|Q_F(\theta)-Q_F\left(\theta^{'}\right)\right| \\
\leq &\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|\left[m_F(\theta,\tau,g)\right]_-^2-\left[m_F\left(\theta^{'},\tau,g\right)\right]_-^2\right| \\
\leq &\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left\{\left|m_F(\theta,\tau,g)-m_F\left(\theta^{'},\tau,g\right)\right|^2+2\left|m_F(\theta,\tau,g)\right|\left|m_F(\theta,\tau,g)-m_F\left(\theta^{'},\tau,g\right)\right|\right\} \\
\leq &\left\{\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F(\theta,\tau,g)-m_F\left(\theta^{'},\tau,g\right)\right|+2\left[\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F(\theta,\tau,g)\right|\right]\right\} \\
&\cdot\left[\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F(\theta,\tau,g)-m_F\left(\theta^{'},\tau,g\right)\right|\right] \\
\leq &C\cdot\left\{C\left[d\left(\theta,\theta^{'}\right)\right]^c+2C\left[d(\theta,\theta_0)\right]^c+2\sup_{(\tau,g)\in\mathcal{T}\times\mathcal{G}}\left|m_F(\theta_0,\tau,g)\right|\right\}\left[d\left(\theta,\theta^{'}\right)\right]^c \\
\leq &B\cdot\left[d\left(\theta,\theta^{'}\right)\right]^c,
\end{aligned}$$

where the second inequality follows from the fact the for all $m_1, m_2 \in \mathbb{R}$,

$$\left|[m_1]_-^2-[m_2]_-^2\right| = \left|[m_1]_-+[m_2]_-\right|\left|[m_1]_--[m_2]_-\right| \leq \left(|m_1-m_2|+2|m_2|\right)|m_1-m_2|,$$

the fourth inequality holds by Assumption 5, the last inequality is a consequence of compactness of Θ and Assumption 4. Condition (iii) is guaranteed by Theorem 2 and definition of $Q_F(\theta)$. Condition (iv) is a consequence of Assumption 6. Finally, Lemma 1 shows condition (v).

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Andrews, D. W. K. (1999): "Estimation when a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

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

Ben-Akiva, M., and B. Boccara (1995): "Discrete choice models with latent choice sets," *International Journal of Research in Marketing*, 12, 9–24.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Berry, S., J. Levinsohn, and A. Pakes (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112, 68–104.

Bronnenberg, B. J., M. W. Kruger, and C. F. Mela (2008): "Database paper-The IRI marketing data set," *Marketing Science*, 27(4), 745–748.

Bruno, H., and N. Vilcassim (2008): "Structural demand estimation with varying product availability," *Marketing Science*, 27, 1126–1131.

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.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 76. Elsevier.

Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80(1), 277–321.

Chernoff, H. (1954): "Rational selection of decision functions," *Econometrica: journal of the Econometric Society*, pp. 422–443.

Chiang, J., S. Chib, and C. Narasimhan (1999): "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics*, 89, 223–248.

Conlon, C. T., and J. H. Mortimer (2008): "Demand estimation under incomplete product availability," Discussion paper, National Bureau of Economic Research.

Crawford, G. S., R. Griffith, and A. Iaria (2016): "Demand Estimation with Unobserved Choice Set Heterogeneity," .

Dominguez, M., and I. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601–1615.

Draganska, M., and D. Klapper (2010): "Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data," Working Paper.

Eliaz, K., and R. Spiegler (2011): "Consideration sets and competitive marketing," *The Review of Economic Studies*, 78(1), 235–262.

Gandhi, A., Z. Lu, and X. Shi (2013): "Estimating Demand for Differentiated Products with Error in Market Shares," .

Gentry, M. (2011): "Displays, Sales, and In-Store Search in Retail Markets," Discussion paper, Working Paper.

Gilbride, T. J., and G. M. Allenby (2004): "A choice model with conjunctive, disjunctive, and compensatory screening rules," *Marketing Science*, 23(3), 391–406.

——— (2006): "Estimating heterogeneous EBA and economic screening rule choice models," *Marketing Science*, 25(5), 494–509.

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

Hauser, J. R., O. Toubia, T. Evgeniou, R. Befurt, and D. Dzyabura (2010): "Disjunctions of conjunctions, cognitive simplicity, and consideration sets," *Journal of Marketing Research*, 47(3), 485–496.

Hauser, J. R., and B. Wernerfelt (1990): "An evaluation cost model of consideration sets," *Journal of consumer research*, pp. 393–408.

Hendel, I., and A. Nevo (2006): "Sales and Consumer Inventory," *Rand Journal of Economics*, 37(3), 543–561.

Hortacsu, A., and C. Syverson (2004): "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds," *The Quarterly Journal of Economics*, 119(2), pp. 403–456.

Jedidi, K., and R. Kohli (2005): "Probabilistic subset-conjunctive models for heterogeneous consumers," *Journal of Marketing Research*, 42(4), 483–494.

Kahn, S., and E. Tamer (2009): "Inference on Randomly Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152, 104–119.

Keane, M. P. (2013): "Panel data discrete choice models of consumer demand," Discussion paper, Economics Group, Nuffield College, University of Oxford.

Liu, Q., and N. Arora (2011): "Efficient choice designs for a consider-then-choose model," *Marketing Science*, 30(2), 321–338.

Manski, C. F., and S. R. Lerman (1977): "The Estimation of Choice Probabilities from Choice Based Samples," *Econometrica*, 45(8), 1977–1988.

Manzini, P., and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82(3), 1153–1176.

Marschak, J. (1960): "Binary-choice constraints and random utility indicators," in *Proceedings of a Symposium on Mathematical Methods in the Social Sciences*, vol. 7, pp. 19–38.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed attention," *The American Economic Review*, 102(5), 2183–2205.

McFadden, D. (1973): "Conditional logit analysis of qualitative choice behavior," .

——— (1978): "Modeling the choice of residential location," *Transportation Research Record*, (673).

Mehta, N., S. Rajiv, and K. Srinivasan (2003): "Price uncertainty and consumer search: A structural model of consideration set formation," *Marketing science*, 22(1), 58–84.

Moon, H. R., and F. Schorfheide (2006): "Boosting Your Instruments: Estimation with Overidentifying InequalityMoment Conditions," IEPR Working Paper No. 06.56, Institute of Economic Policy Research (IEPR).

Moraga-González, J., M. Wildenbeest, and Z. Sandor (2010): "On the identification of the costs of simultaneous search," .

Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2009): "Consumer search and prices in the automobile market," .

Musalem, A., M. Olivares, E. Bradlow, C. Terwiesch, and D. Corsten (2010): "Structural Estimation of the Effect of Out-of-Stocks," *Management Science*, 56(7), 1180–1197.

Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Paola, M., and M. Marco (2013): "Stochastic Choice and Consideration Sets," .

Pires, T. (2012): "Consideration Sets in Storable Goods Markets," *working paper, Northwestern University.*

Powell, J. L. (1984): "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 25(3), 303–325.

Roberts, J. H., and J. M. Lattin (1991): "Development and Testing of a Model of Consideration Set Composition.," *Journal of Marketing Research (JMR)*, 28(4).

Santos, B. D. L., A. Hortacsu, and M. R. Wildenbeest (2012): "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 106, 2955–2980.

Sen, A. K. (1971): "Choice functions and revealed preference," *The Review of Economic Studies*, pp. 307–317.

Shi, X. (2015): "Model selection tests for moment inequality models," *Journal of Econometrics*, 187, 1–17.

Shocker, A. D., M. Ben-Akiva, B. Boccara, and P. Nedungadi (1991): "Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions," *Marketing letters*, 2(3), 181–197.

Train, K. E. (2009): *Discrete choice methods with simulation*. Cambridge university press.

Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010): "Retrieving unobserved consideration sets from household panel data," *Journal of Marketing Research*, 47(1), 63–74.