JOURNAL OF Econometrics

CONTENTS

Robust firm pricing with panel data

Benjamin R. Handel [a,b], Kanishka Misra [c,d], James W. Roberts [e,b,*]

[a] Department of Economics, University of California at Berkeley, United States

[b] NBER, United States

[c] London Business School, United Kingdom

[d] The Ross School of Business, University of Michigan, United States

[e] Department of Economics, Duke University, United States

ABSTRACT

Firms often have imperfect information about demand for their products. We develop an integrated econometric and theoretical framework to model firm demand assessment and subsequent pricing decisions with limited information. We introduce a panel data discrete choice model whose realistic assumptions about consumer behavior deliver partially identified preferences and thus generate ambiguity in the firm pricing problem. We use the minimax-regret criterion as a decision-making rule for firms facing this ambiguity. We illustrate the framework's benefits relative to the most common discrete choice analysis approach through simulations and empirical examples with field data.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Standard approaches for applying random utility models to interpret discrete choice data maintain assumptions that allow point identification of consumer preferences, which eases counterfactual choice predictions. The point of this paper is to relax our modeling assumptions of consumer behavior without compromising our ability to provide guidance on counterfactual outcomes. To this end, the first part of this paper develops a model that requires only conservative assumptions about consumer decision-making processes to partially identify preferences and, consequently, counterfactual choices. We focus on settings with panel data, and extend prior work by integrating conservative assumptions on inter-temporal decision-making into our econometric framework.

The second part of the paper focuses on how firms use the model's output to make strategic decisions. While our robust modeling assumptions still allow us to predict counterfactual choices, the analysis of a firm's strategic optimization problem is complicated by the fact that those counterfactual demand curves are only partially identified. With this information set, how does a firm make a strategic choice like a pricing decision? When output is point identified, a firm has complete information about the distribution of consumer preferences, and can use this information to maximize expected profits. In our setting, and indeed in any setting where a firm uses partially identified parameters as decision-making inputs, the firm may not be able to construct a prior over the set of feasible preference parameters in order to maximize expected profits. In this sense the firm faces "Knightian Uncertainty", or ambiguity, about consumer demand. The second primary contribution of this paper analyses how a firm can use the partially identified set of preference distributions arising from our conservative econometric model to choose prices under ambiguity.

This paper thus integrates the prior theoretical work on firm pricing under ambiguity with a novel econometric framework to (i) econometrically model the lack of information inherent in the firm's problem when only conservative assumptions about consumer decision-making are made and then to (ii) study how firms will price under ambiguity if their information set is

* Correspondence to: 213 Social Sciences Building, 419 Chapel Drive, Durham, NC 27708, United States. Tel.: +1 919 660 1822.

E-mail addresses: handel@berkeley.edu (B.R. Handel), kmisra@london.edu (K. Misra), j.roberts@duke.edu (J.W. Roberts).

consistent with the output of our econometric model. We then use this two part framework to study firm pricing in both simulations and field data. We show that there are many cases where, despite its more conservative approach, our integrated model compares favorably to, and at times outperforms, the combined mixed logit and expected profit maximization framework, which is the "workhorse" model of the industrial organization literature.

We investigate an environment with panel data and develop four alternative models that correspond to different assumptions on how consumer preferences can change over time. Each model results in a distinct, partially identified set of consumer preferences and, consequently, demand curves and counterfactual choices. Across these models, the primary parametric assumption we maintain is that consumer preferences are a linear function of product attributes, as in the canonical discrete choice framework of McFadden (1974). Under this assumption, each alternative inter-temporal decision-making framework places restrictions on the range of feasible valuations for products and their associated attributes, given consumers' choices. Unlike models with more powerful statistical assumptions about the distribution of preferences (e.g. random coefficients) and the distribution of idiosyncratic preference shocks (e.g independent and identically distributed logit errors), each of our models can be rejected by the data if the underlying assumptions are violated, increasing the credibility of the analysis at the expense of reduced precision.

The four frameworks differ according to their maintained assumptions on the time variation of consumers' preferences.[1] In the first, most basic, setup, consumers have the same exact preferences in each time period, with no idiosyncratic component, and we directly apply the strong axiom of revealed preferences to partially identify consumer preferences. This most basic model lacks flexibility in allowing for within-consumer variation in preferences over time, and hence will likely be rejected by the data. Thus we extend it in three ways to allow for time varying consumer utility. First, we study random shocks to each consumer's utility for each product and time period. Unlike previous models in this literature, which place more structure on the distribution of these random shocks, we only maintain that these errors are bounded in size by a constant in absolute value and we do not make any independence or distributional assumptions about these error terms.[2] Including bounded errors allows the model to account for small departures from stable preferences that occur often over the course of multiple decisions. We illustrate how the bound (i) is identified and (ii) can be estimated in a first stage using only the original panel data set.[3] Our next framework studies data contamination, an oft-cited determinant of observed time-variation in purchases (see e.g. Keane (1997) or Einav et al. (2010)). Intuitively, this allows the model to account for large departures from stable preferences that occur rarely. It will often be the case that for particular data sets there is existing knowledge that can be drawn upon to inform the econometrician about the extent of contamination in the data. Our final framework combines our analysis of bounded errors with our analysis of data contamination.

With these partially identified predictions in hand, we then investigate the firm pricing problem under ambiguity. To our knowledge, this is the first work that integrates an econometric

framework that generates partially identified demand, due to a firm's lack of information on consumer preferences, with a model of firm pricing under ambiguity. We model firm decision-making using the minimax-regret pricing criteria discussed elsewhere in a purely theoretical setting (see e.g. Bergemann and Morris (2005) Bergemann and Schlag (2007) or Bergemann and Schlag (2008)). Under this criterion, the firm chooses a price to minimize its maximum *regret* over the set of perceived feasible demand curves represented by the partially identified output of our econometric model. Here *regret* is defined for a given demand curve in the set of feasible demand curves, and equals the difference between profits under the *optimal* price for that demand curve and the profits under the *actual* price.[4] Our analysis considers the cases of monopolistic and duopolistic pricing under ambiguity based on the partially identified set of demand curves where the latter incorporates a strategic environment.[5] While we use minimax-regret as a criterion because it has the desirable property that it trades off potential losses from overpricing (selling too little) versus those from underpricing (not extracting enough consumer value), we note that any criterion for decision-making under ambiguity could be used to make decisions with our partially identified econometric output. For example, the maxmin criterion (see e.g. Gilboa and Schmeidler (1989)) is a potential alternative to minimax-regret, which we study briefly in the context of our empirical examples.

We use simulations to test the performance of our joint econometric–theoretical framework relative to two benchmark specifications: (i) the mixed logit with multivariate mixing and (ii) *ex post* optimal pricing under perfect information. For plausible underlying data generating processes, we analyze how consumer choices translate into partially identified estimates of demand for our different econometric models. The results show that, in the monopoly setting, the monopolist gets *close* to *ex post* efficient prices with our framework independent of the underlying error data generating process. On the other hand, the mixed logit performs well if the underlying data structure has i.i.d. errors but can yield large differences from optimal pricing when this is violated (such as when there are time correlated error shocks). In the oligopolistic setting, we analyze minimax-regret best response curves given partially identified preferences and show that prices under our model are much closer to the *ex post* efficient prices for many data generating processes. These results suggest that our integrated framework for robust firm pricing provides a viable alternative to the canonical mixed logit model in cases where it is likely that the firms studied have limited information.

Finally, we illustrate how our methodology can be applied in an actual empirical setting in order to recommend an optimal price when only conservative assumptions about consumer decision-making are made. The setting we consider is retailer milk pricing. Fluid milk is a frequently purchased non-storable good and is an important category for retailers as it has the highest penetration of any retail category (Bronnenberg et al. (2008)). It is mainly driven by retailer owned, private label brands and, importantly for us, it is a non-storable good.[6,7] By first estimating demand

---

[1] As an additional extension to each of these four frameworks, we develop a method to use cross-sectional variation in conjunction with each of these inter-temporal frameworks to obtain further identifying power when the panel is not a representative sample from the population.

[2] These assumptions address many of the undesirable features of the standard extreme value random utility model, as discussed in Bajari and Benkard (2003).

[3] We characterize the identified sets in the cases (i) where the error bound is known by the firm and econometrician and (ii) where the error bound is unknown by both parties and is estimated.

[4] Note that this notion of regret from the statistical decision literature (e.g. Savage (1951)) is completely distinct from the notion of regret discussed in the psychology and economics literature.

[5] We developed our framework for strategic firm pricing under ambiguity simultaneously and independently of recent work by Renou and Schlag (2010) who study foundations for minimax-regret strategic pricing equilibrium in a purely theoretical paper.

[6] See Fong et al. (2011) for a review of marketing papers estimating private label elasticity with standard models.

[7] This second observation is relevant as this rules out stockpiling behavior which will generate dynamic choice behavior, the modeling of which lies outside the scope of this paper (for papers that do model this behavior based on more "traditional" demand estimation techniques, see Erdem et al. (2003) or Hendel and Nevo (2006)).

and then solving for minimax-regret optimal prices, we show that our methodology is applicable in real-world settings and returns sensible counterfactual recommendations (the minimax-regret optimal price is $2.40/gallon when actual observed prices average about $2.56/gallon).

This paper helps to advance the twin goals in the broader discrete choice literature of (i) describing preferences and (ii) making counterfactual predictions. Papers that best describe preferences in specific contexts can make conservative assumptions with simple decision-theoretic foundations, but, as a result, are generally not well suited for counterfactual prediction. For example, Samuelson (1938, 1948) study observed consumers' choices from different choice sets and price–income pairs and use either the weak or strong axiom of revealed preference along with a transitivity assumption to draw powerful conclusions about preferences *for products in the observed environment*. Following this line of work, Varian (1982, 1983) develop an econometric methodology that (i) tests if observed choice behavior is consistent with rational choices and (ii) recovers preferences as a function of prices and budget sets. While these approaches infer preferences under minimal assumptions, their empirical viability is limited because they require very rich choice data, the models can be easily rejected by the data, and they cannot inform predictions in counterfactual settings.

In more recent work, Blundell et al. (2008) use revealed preference restrictions to non-parametrically identify demand responses along Engel curves.[8] Similar to this paper, their objective is to use theoretical restrictions to obtain credible preference estimates, given individual-level data on relative prices and total expenditures, without imposing the usual parametric and statistical assumptions that permeate the demand estimation literature. Our approaches differ along multiple dimensions, most prominently that Blundell et al. (i) take a cross-sectional approach that does not incorporate time series decision restrictions in identification and (ii) do not incorporate the notion of products as bundles of attributes (or maintain the linear utility in attributes assumption that we do). The former implies that our primary sources of identification and ideal data sets differ substantially, as we incorporate theoretical restrictions on how a given consumer makes decisions over time. While Blundell et al. study what minimum level of statistical perturbation to their consumer utility bounds can justify a rational paradigm given their model and choice environment, our analysis partially identifies sets of preferences for attributes and can be used to make predictions in counterfactual choice settings where products are composed of those same attributes (e.g. with new products or new choice settings based on existing products). Finally, we incorporate the possibility of data contamination and link the output of the econometric exercise directly to the firm pricing decision under ambiguity.

The second, and much more heavily utilized, branch of the discrete choice literature makes stronger assumptions about consumer behavior but is also able to make stronger statements about counterfactual outcomes. These papers assume that consumers have preferences for product attributes which are aggregated to establish preferences over products (see Lancaster (1966) and McFadden (1974)). The canonical model assumes that consumer preferences have a specific parametric form that maps attribute-specific preference parameters, vectors of product attributes, and an additive preference shock known to the firm (but not the researcher) into product values and preference orderings. Papers in this literature make different assumptions about the distributions of deterministic preference heterogeneity and the idiosyncratic preference shock. Many assume that the error terms are either independent and/or identically distributed across consumers, products and time.[9] Given the distribution of error shocks and form of utility, model parameters are identified using observed choice data.[10] Our work uses some of the basic assumptions in this literature, such as attribute based preferences, to maintain the ability to perform flexible counterfactual analyses, but refrains from making parametric assumptions which are without theoretical foundation and can be difficult to interpret.[11]

Of special note is recent work by Manski (2007), who studies a semi-parametric cross-sectional discrete choice model with no assumptions on the distribution of errors. Manski partially identifies preferences based on three main sources: (i) linear utility in attributes, (ii) the consistency of preference parameters with observed rational behavior for a given choice set and (iii) cross-sectional variation in prices and choice sets. Our paper uses (i) but has different notions of (ii) and (iii), primarily because we study a panel setting where choice consistency over time must be taken into. We view his paper as complementary to our own from an econometric perspective since it has a similar underlying motivation but applies to a distinctly different data environment. Moreover, his paper does not investigate how the output of the model will be used in decision-making as we do with our emphasis on firm pricing.

The remainder of the paper proceeds as follows. Section 2 sets up the model and derives the identification regions for each framework. Section 3 discusses estimation. Section 4 describes the firm problem when preferences are partially identified. Section 5 illustrates the methodology through simulations and in an application to milk pricing. Section 6 concludes.

2. Model

The problem we consider is one where a firm observes panel data and uses them to make a pricing decision. Our goal is to relax the assumptions underlying the traditional literature on such behavior by not requiring the firm to know the distribution of demand for its products. That is, we consider a firm which seeks

---

[8] In related work Blundell et al. (2003) show how to use non-parametric methods to detect revealed preference violations.

[9] It is important to note that the mixed-logit class of models does allow for error correlations (see e.g., McFadden and Train (2000)). Several other papers provide counterexamples to this claim and deserve specific mention. An important contribution in this line of work is Keane (1997) who establishes the presence of state dependent preferences as well as heterogeneity in these tastes. Recently, Fiebig et al. (2010) extend the mixed multinomial logit model to present a generalized multinomial logit model that allows better modeling of consumers with extreme and/or random tastes (in the sense that a particular attribute of a product drives much of their decision-making). In other related work, Geweke (2012) explores recovering regions of parameters based on observed data but from a Bayesian perspective.

[10] Broadly speaking, these models fall into four categories (Ben-Avika et al. (1997)): those assuming (1) functional forms for deterministic utility (linear in product attributes) and that error terms are i.i.d. according to a specified distribution, such as Type 1 Extreme Value (this could include dynamic structural models of demand, e.g. Erdem and Keane (1996), Erdem et al. (2003) or Hendel and Nevo (2006)); (2) a parametric functional form for deterministic utility (usually linear in attributes) but with unspecified error distribution (see Manski (1975)); (3) a specific form for the error distribution, but no functional form assumption on deterministic utility (see Haistie and Tibshirani (1990) and Abe (1995)); and (4) no functional form for deterministic utility or the distribution of error terms (see Matzkin (1993)).

[11] We view our work as complementary to the prior literature that maintains stronger assumptions on the distributions of preferences and preference shocks. If the output from a model with many maintained assumptions does not lie within the bounds our models produce, the researcher should be skeptical that their model is correctly specified. Further, if the point identified output lies near one edge of our feasible demand curve set, our model sheds light on the likely direction of any potential model bias. Finally, if the researcher believes there are specific justifications for the parametric assumptions maintained, this adds insight above and beyond our model.

to maximize profits, but cannot necessarily do so in the traditional way because it does not know the distribution of types in the population. This implies that the firm can only partially identify demand for each potential price it considers. In this section we show how to non-parametrically identify consumer preferences using panel data. We consider a variety of models with increasing flexibility to illustrate how one can identify demand under specific incremental assumptions.

2.1. Base case: time consistency

We begin our analysis with the base model that assumes that each individual has stable preferences over time. There are no product or time specific preference shocks, which yields tight bounds, but a high probability that the model will be rejected when consumers' decisions cannot be rationalized within the linear preferences over attributes specification.

2.1.1. Model 1: time consistent preferences

As in Manski (2007) we examine the discrete choice problem faced by consumers within a treatment response framework. In this setup, there are A possible distinct alternatives (products) each uniquely characterized by a K dimensional attribute vector x. Each attribute is assumed to have finite support,[12] therefore the set A is finite. We define the set of possible treatments D as the space of possible choice sets an individual could face, which in this setting is the collection of all non-empty finite subsets of A. Each individual faces a choice set from D and responds by choosing an element of that set. Formally, there is a population of N individuals, denoted $\mathcal{I}$, in which each individual $i \in \mathcal{I}$ has a response function $y_i(\cdot) : D \rightarrow A$ mapping choice sets into unique choices from that set. The probability distribution $P[y(\cdot)]$ of the random function $y(\cdot) : D \rightarrow A$ describes the aggregate choices (product shares) made by the population.

For example, consider a case where there are three feasible alternatives b, c, and d, so that $A = \{b, c, d\}$. Assume that the alternatives are described only by their identified name (b, c or d). In our notation we would say that $K = 3$, as each alternative is described by three indicators (similar to fixed effects). Suppose that we observe data from a choice setting where N consumers choose between product b and c, so that $D = \{b, c\}$. In our notation we would say that $y_i(\{b, c\}) = b$ for the N_b consumers who choose b, $y_i(\{b, c\}) = c$ for the N_c consumers who choose c and that $P[y(\{b, c\}) = b] = \frac{N_b}{N}, P[y(\{b, c\}) = c] = \frac{N_c}{N}$.

Our objective is to estimate counterfactual choice probabilities. For example, what percentage of consumers would choose d in a choice between all three alternatives? Without any assumptions about the underlying consumer decision-making process, we cannot say anything informative about this counterfactual. In our setting we make several conservative assumptions on individual behavior that will allow us to make counterfactual predictions.

Assumption 1 (*Utility Maximization*). Consumers have well defined preferences and make decisions that maximize utility subject to the available elements in their choice set.

Under this assumption, if consumer i faces choice set D we have the following information about the consumers response function where $u_{i,a}$ is the utility consumer i gets from alternative a:

$$y_i(D) = \underset{a \in D}{\operatorname{argmax}}\ u_{i,a}. \tag{1}$$

This assumption allows us to make inferences about counterfactual consumer choices, and we can classify the population into types based on their preferences. A type is defined by preferences over all elements of D. There are $|D|!$ possible types in the population corresponding to different permutations of the elements of D that could correspond to rational preference orderings. Inference about what type a consumer might be can be made from observed choice data. In the simple example above, there are 6 types of consumers:

1. $b \succ c \succ d$,   2. $b \succ d \succ c$,   3. $c \succ b \succ d$
4. $c \succ d \succ b$,   5. $d \succ b \succ c$,   6. $d \succ c \succ b$.

Observing the fact that N_b consumers choose option b in a choice between $\{b, c\}$ implies that these consumers are of type 1, 2 or 5. Similarly, observing N_c consumers choose c implies that they are of type 3, 4, or 6. In counterfactual choice settings, the proportion of consumers who choose c from a choice set $\{b, c, d\}$ is equivalent to estimating the proportion of consumers of type 3 or 4. In this example this is bounded above at $\frac{N_c}{N}$. However, without making more assumptions about the underlying utility structure, we cannot estimate the counterfactual choice probability for choice d since we never observe d in the consumer choice set.

Therefore, we follow the discrete choice literature (McFadden (1974)) and consider products to be bundles of attributes and assume that consumers' utility functions are linear in these attributes.

Assumption 2 (*Linear Utility*). Individual utility functions are linear in the K dimensional attribute vector x, that describes the alternatives in the choice sets.

Additionally, we define individual specific K dimensional parameters ω_i to describe individual i's preferences for each attribute. We define Ω as the feasible parameter space for these preferences, with $\omega_i \in \Omega$.

Under these characterizations, the utility consumer i gets from alternative a is defined as:

$$u_{i,a} = \omega_i \cdot x_a. \tag{2}$$

Relating this to the response function we now have:

$$y_i(D) = \underset{a \in D}{\operatorname{argmax}}\ \omega_i \cdot x_a. \tag{3}$$

In our formulation of this utility model, price is one attribute in x_a. Thus, we consider one product sold at P different prices as P distinct feasible alternatives in the set of all alternatives A. Here, a demand curve involves constructing a set of counterfactual predictions based on the set of alternatives.

Unlike our simple example where we had six types of consumers in our model with linear utility, we now have a continuous parameter space with infinite consumer types. However, since we have a space of finite alternatives, we can represent the continuous space Ω by a discrete distribution of types corresponding to the different possible choice functions over A (as in Manski (2007)). These representations are equivalent because each of the $\omega \in \Omega$ that corresponds to the same preference ordering over all alternatives cannot be identified from each other in the data in our model. Formally, let A_m, $m = 1, \ldots, |A|!$ represent the mth permutation of A. If $x_{m,n}$ is the attribute bundle of the nth element of A_m, then the discrete type space can be defined:

$$\Theta_m \equiv [\omega \in \Omega : \omega \cdot x_{m1} > \omega \cdot x_{m2} > \cdots > \omega \cdot x_{m|A|}].$$

Let θ_l denote a generic element of Θ_l which we can use from this point forward to represent that type without loss of generality.

---

[12] This assumption contrasts with those in Berry and Haile (2009).

We have now mapped our parameter space from a continuous set Ω to a set of discrete types $\{\theta_1, \ldots, \theta_{|A|!}\}$.[13]

We partially identify the distribution of types with panel data by using this model to identify the feasible range of preferences for each individual based on their choices and then aggregating these to form an aggregate bound on the distribution of types. In our panel data, for each individual $i \in \mathcal{I}$ we observe choices a_{it} from distinct choice sets over time (e.g., choice made every week) d_{it} for $t = 1, \ldots, T_i$.[14]

The main advantage of these data are that we can use all T_i observations for the individual to gain more information about a given individual and aggregate feasible types. However, the panel framework also presents additional complications since it is possible for data on individual decisions to be inconsistent with a constant preference parameter over time. Below we present models that allow for the most commonly given explanations for such apparent inconsistencies: time varying preferences and data contamination.[15] We begin, however, with a base model that assumes that consumers have stable preferences over time and that there is no data contamination. It is the strictest of the models we present since it is the least flexible in terms of how it can rationalize a sequence of observed choices.

Definition 1 (*Time-Consistent Utility*)**.** An individual in the population is time-consistent if they always make decisions according to a fixed θ_i.

This definition implies that a consumer's utility in each purchase occasion is described completely by observable (to the researcher) attributes. An individual can be time-consistent if:

$$\Theta_i^1 \equiv \left\{ \theta : \theta \in \cap_{T_i} \{\theta : \theta \cdot x_{a_{it}^*} > \theta \cdot x_{a_{it}^{-*}}, \forall a_{it}^{-*} \in d_{it}\} \right\} \neq \emptyset$$

where a_{it}^* is i's purchase decision at time t and a_{it}^{-*} is an element not chosen from that set. Here, Θ_i^1 denotes the set of feasible θ for individual i given the decisions we observe over time. Under the time-consistency assumption, the partially identified probability of an individual i being of type θ is:

$$H[\Pr(\theta_i = \theta)] = [\Pr(\Theta_i^1 = \{\theta\}), \Pr(\theta \in \Theta^1)]. \tag{4}$$

This says that the lower bound of a specific consumer being a certain type is the probability that the identification region for an individual includes only that type, while the upper bound is the probability that a given type is included in an identification region. If an individual is not time-consistent so that $\Theta_i^1 = \emptyset$, we conclude that individual is not in the $|A|!$ rational types and lies instead in the larger collection of types that is described by all permutations of possible choices across feasible choice sets.

In our base model, we consider the sample that we analyze to be the population of interest. This simplifies exposition of our model in that it allows us to focus on identification instead of sampling properties at this point. The set of feasible *distributions* for θ in the population comes directly from identifying the feasible types

at the individual level (as described above) and then determining all combinations of these types aggregated to the population level. Then, all such feasible aggregations describe the partially identified set of type distributions. More formally, the set of feasible distributions satisfies:

$$H[F(\theta)] \equiv \left\{ F(\theta) | f(\theta) = \frac{1}{N} \sum_{i \in \mathcal{I}} \mathbf{I}[\theta = \theta_i], \forall i \, \forall \theta_i \in \Theta_i^1 \right\}. \tag{5}$$

To understand the set definition, consider the example from above where there are six consumer types. Suppose that there are two consumers for whom, based on their sequence of purchases, we have determined that the first can be type 1 or 2 and the second can be type 2, 3 or 4. Then there are six possible distributions and $H[F(\theta)]$ includes all six possibilities:

1. $f(1) = 1/2, f(2) = 1/2;$ 2. $f(1) = 1/2, f(2) = 1/2$
3. $f(1) = 1/2, f(4) = 1/2;$ 4. $f(2) = 1$
5. $f(2) = 1/2, f(3) = 1/2;$ 6. $f(2) = 1/2, f(4) = 1/2.$

The knowledge of the partially identified distribution of preferences allows us to study counterfactual choice settings and, hence, counterfactual demand. At the individual-level, a given consumer either could or could not choose product a from counterfactual choice set D. This binary possibility depends directly on whether the partially identified preference set for that individual contains *at least one* preference profile where the individual would choose a from D. Given this, we define bounds on demand for product a when the population faces choice set D relative to $H[F(\theta)]$, the set of feasible preference distributions at the population level. Minimum demand for product a comes from the feasible distribution in $H[F(\theta)]$ where the fewest consumers would purchase a (and vice-versa for the maximum):

$$H[P(y(D)) = a] = \left[\min_{F(\cdot) \in H[F(\cdot)]} \sum_\theta \mathbf{1}[y(D) = a] f(\theta), \right.$$
$$\left. \max_{F(\cdot) \in H[F(\cdot)]} \sum_\theta \mathbf{1}[y(D) = a] f(\theta) \right]. \tag{6}$$

2.2. Relaxing time consistency: bounded preference shocks and data contamination

The base model above is predicated on consumers having constant preferences over time. Given this inflexibility, it is possible that an individual's observed purchase decisions cannot be rationalized by such a model. Therefore, we now present two extensions to make the model more realistic.

The first of these allows for individual-time-product specific preference shocks as most discrete choice models do, but assumes no structure on the population distribution of the shocks except that they are bounded. This approach is motivated by Bajari and Benkard (2003) who illustrate that canonical discrete choice models with unbounded errors have some notable undesirable properties. First, as the number of products in the choice set becomes large, the standard approach implies that all consumer decisions are driven by unobserved error shocks. This implies that in settings with large choice sets researchers cannot learn about underlying consumer preferences. Second, in any choice setting, the standard approach implies that every product has a non-zero probability of being chosen by a given consumer, regardless of underlying preferences. In this section, our model with bounded errors allows researchers to learn about preferences over time even with a large number of products and, additionally, allows for a zero probability of choosing a dominated product.

The second of these extensions allows for the possibility that the data may be contaminated and thus observed purchases do

---

[13] As shown in Manski (2007) the linear utility specification has some identifying power as we reduce the number of feasible choice functions before going to the data. In the simulations that we study in Section 5.1, the dimensionality of feasible discrete types is reduced approximately by a factor of ten when we impose the linear model, implying that the number of points in the distribution that we are estimating is also reduced by a factor of ten.

[14] For simplicity, we will consider models where consumers make only one discrete choice at each point in time, though nothing about our setup precludes us from observing multiple choices at multiple points in time for a given individual, where linking contemporaneous decisions would also add identifying power.

[15] Two other potential explanations for such inconsistencies are decision-making errors and non-linear utility. To focus on the core issue of robust firm pricing, we leave the exploration of these types of models to future work.

not reflect actual choices. There are numerous reasons that data may be contaminated including, but not limited to, recording errors, non-response, or interpolation/extrapolation. Heuristically, bounded errors allow the model to explain consumers making frequent, but small departures from stable preferences. Data contamination, on the other hand, permits less frequent, though larger fluctuations in implied preferences.

2.2.1. Model 2: random utility

Most attempts to estimate demand with panel data employ a utility model of the form:

$$u_{i,a} = \theta_i \cdot x_a + \varepsilon_{i,a}. \tag{7}$$

A consumer's utility in each purchase occasion is described by observable (to the researcher) attributes and unobserved error shocks. Here we make the following assumption on ε:

Assumption 3 (*Random Utility Model with Bounded Errors*)**.** An individual in the population receives random utility shocks $\varepsilon_{i,a,t}$ for each i, a, and t. The only assumption about these shocks are that they are (strictly) bounded within some range $[-\delta, +\delta]$.

Therefore, at each point in time, and for any product, an individual receives a shock to his utility of magnitude no greater than δ. Further, as opposed to what is done in the literature, we make no distributional or independence assumptions about this shock, apart from setting this bound.[16] In theory, the firm could know δ or calibrate it from a variety of data sources. In Section 3 we present one methodology for how a firm may non-parametrically select an appropriate δ in a given empirical setting.

Without any assumptions about the distribution of ε, all we know is that $\forall a_1, a_2 \in D, -2\delta \leq \varepsilon_{i,a_1,t} - \varepsilon_{i,a_2,t} \leq 2\delta$. Therefore the identification region for this model is given:

$$\Theta_i^2 \equiv \{\theta : \theta \in \cap_{T_i} \{\theta : \theta \cdot x_{a_{it}^*} \geq \theta \cdot x_{a_t^{-*}} - 2\delta, \forall a_t^{-*} \in d_{it}\}\}. \tag{8}$$

As in the time-consistency case, the partially identified probability of an individual i being of type θ and the feasible population distributions of types are given by Eqs. (4) and (5), respectively, with Θ_i^2 replacing Θ_i^1.

As in the base model, knowledge of the partially identified distribution of preferences allows us to study counterfactual choice settings. We can derive the probability that alternative a is chosen from a choice set D as was done in Eq. (6).

2.2.2. Model 3: data contamination

While bounded errors provide flexibility in terms of describing consumer decisions that depart by a small magnitude from their stable preferences, in some cases there may be large fluctuations in implied preferences for an individual as a function of observed choices. Here, we expand the base model by allowing for a proportion of the data to be contaminated. An alternative interpretation of large deviations from apparently stable preferences that is based more on behavioral foundations, is that consumers occasionally make sub-optimal decisions.[17] We prefer

the data contamination explanation and so proceed under that interpretation.[18, 19]

Assumption 4 (*Data Contamination Model with Time-Consistent Utility*)**.** Individuals have time-consistent utility but up to ϕ percentage of their decisions could be recorded with error.

There are numerous reasons that data may be contaminated and multiple papers have explored the extent of data contamination and issues relating to identification and estimation in its presence (see for example Horowitz and Manski (1995), Keane (1997), Erdem et al. (1999) or Einav et al. (2010)). Define T_i^ϕ as the set of observations for i that are not contaminated, with $T_i^\phi \subseteq T_i$ and $|T_i^\phi| \geq (1-\phi)|T_i|$. Given T_i^ϕ the partially identified set of preferences for each consumer are defined[20]:

$$\Theta_i^3(T_i^\phi) \equiv \left\{\theta : \theta \in \cap_{T_i^\phi} \{\theta : \theta \cdot x_{a_{it}^*} \geq \theta \cdot x_{a_t^{-*}} \forall a_t^{-*} \in d_{it}\}\right\}. \tag{9}$$

For a given bound ϕ, there are many possible candidates for T_i^ϕ since the researcher cannot observe which data are actually contaminated. As a result, all feasible T_i^ϕ must be considered to determine the true partially identified set Θ_i^3:

$$\Theta_i^3 \equiv \cup_{T_i^{\phi*}} \Theta_i^3(T_i^\phi). \tag{10}$$

Here, $T_i^{\phi*}$ is the set of all feasible T_i^ϕ. The union of these sets is the correct metric, because each potential T_i^ϕ could reflect the set of all actual purchases. It is important to note that, from an empirical standpoint, the only potential T_i^ϕ the researcher needs to consider are those that have exactly $1 - \phi$ observations ($|T_i^\phi| = (1 - \phi)|T_i|$). This is true because the set of potential preferences can only expand as restrictions from choices are removed, so those feasible data sets with the fewest choice restrictions (largest possible proportion of misclassified data) will be the most inclusive.

As in the time-consistency case, the partially identified probability of an individual i being of type θ and the feasible population distributions of types are given by Eqs. (4) and (5), respectively, with Θ_i^3 replacing Θ_i^1.[21] Counterfactually, once we determine the partially identified distribution of preferences we can derive the probability that alternative a is chosen from a choice set D as was done in Eq. (6).

---

[16] If there are distinct consumer types in the data which are observable, then δ could vary across consumer types. For simplicity we assume no such distinctions exist as the extension is straightforward.

[17] This may occur for a variety of reasons. It may be, for instance, that consumers do not have full information about the options available and make decisions based on some individual heuristic. For example, there is evidence showing that time-constrained consumers are more likely to purchase items from the middle of store shelves (Dreze et al. (1994)). As researchers, we do not know when individuals use such heuristics, nor which heuristics they use.

[18] There is an added benefit of avoiding a model of sub-optimal decision-making. If consumer decisions are actually random some fraction of the time, a firm may wish to set an infinite price. As we study firm pricing below, it would be difficult to rationalize this strategy with empirical evidence that prices are rarely infinite. We thank the referees for pointing this out.

[19] An alternative interpretation of ϕ could be that with a small probability, a particular consumer's observation is recorded with error. We do not adopt this approach because if ϕ were the probability that an individual observation was misclassified, then it is possible, albeit unlikely, that all observations in any one data set would be contaminated. This would be equivalent to having our interpretation of ϕ as a proportion and that proportion being 1. ϕ can be thought of as a value that will be above the proportion of misclassified data for almost every individual.

[20] A related process is discussed in Keane and Sauer (2009, 2010), but in a different context where the authors consider the case of employment status misclassification when modeling female labor supply.

[21] We assume that at most ϕ percent of observations are misclassified/contaminated for each individual. If the true data generating process (DGP) is binomial where, with some probability p, an observation is recorded with error, any bound less than $\phi = 1$ could in theory be violated. In the context of our assumed bound, if a consumer actually has a proportion of misclassified data greater than ϕ, the model will be rejected by the data if the misclassification, treated as an actual choice, leads to implied preferences that are not consistent across purchases. In this case, the identified set of parameters will be empty. Our model already conservatively accounts for the case where actual misclassifications are less than the assumed bound ϕ.

2.2.3. Model 4: random utility with data contamination

Assumption 5 (*Data Contamination Model with Random Utility*)**.** Individuals have random utility with bounded errors and up to ϕ percentage of their decisions could be recorded with error.

This model is a combination of models 2 and 3 discussed above. There are two parameters: δ, the bound for the random utility shocks and ϕ, the bound for the frequency of data contamination. The combination of these factors is attractive in situations where there is a large probability that individual preferences change by small amounts over different choice settings and a small probability that an individual appears to makes a decision that departs completely from our description of their preferences.

To estimate consumer preferences with these assumptions given a feasible T_i^ϕ, we define:

$$\Theta_i^4(T_i^\phi) \equiv \left\{ \theta : \theta \in \cap_{T_i^\phi} \{ \theta : \theta \cdot x_{a_{it}^*} \geq \theta \cdot x_{a_{it}^{-*}} - 2\delta \ \forall a_t^{-*} \in d_{it} \} \right\} \quad (11)$$

where T_i^ϕ is defined as above. Then, as in model 3, the full partially identified set of consumers preferences is defined:

$$\Theta_i^4 \equiv \cup_{T_i^{\phi*}} \Theta_i^4(T_i^\phi). \quad (12)$$

As in the time-consistency case, the partially identified probability of an individual i being of type θ and the feasible population distributions of types are given by Eqs. (4) and (5), respectively, with Θ_i^4 replacing Θ_i^1.[22] Counterfactually, once we determine the partially identified distribution of preferences we can derive the probability that alternative a is chosen from a choice set D as was done in Eq. (6). As with model 2, the firm could know δ or calibrate it from a variety of data sources. Section 3 presents one methodology for how a firm may non-parametrically select an appropriate δ conditional on ϕ. In our empirical exercise in Section 5, we calibrate ϕ using prior work on data contamination, and discuss several alternative approaches.

As an additional extension to each of these four frameworks, in Appendix A we develop a method to use cross-sectional variation in conjunction with each of these inter-temporal frameworks to obtain further identifying power when the panel is not a representative sample from the population.

2.2.4. Price endogeneity and bounded errors

While most discrete choice models rely on exogenous variation of the independent variables, such as price, our model makes no explicit independence assumptions and therefore does not require exogenous variation. In general, endogeneity is a much bigger concern with aggregate purchase data (see e.g. Berry et al. (1995)) than when the researcher has panel scanner data in our setting (see e.g. Erdem et al. (2003)). While endogeneity is thus not likely to be a major concern in our context, we note as a robustness point that the partially identified estimates of consumer preference distributions would not be biased *even if* firms had any additional amount of information that they were incorporating into prices.

To see this consider the utility function as in Berry et al. (1995) a canonical model in industrial organization, where $u_{ijt} =$ $\beta_i X_{jt} + \xi_{jt} + \varepsilon_{ijt}$ where ξ_{jt} is a common shock that impacts all consumers. The common assumption is that the firm observes ξ_{jt} and therefore sets a price P_{jt} that includes this information. In our specifications with bounded errors, we consider ξ_{jt} to be a part of ε_{ijt}, which, in most standard discrete choice models with independent and i.i.d. errors would cause price endogeneity as the error term would be correlated with the independent variables. However, in our model we make no independence or identical distribution assumptions and therefore, with the caveat that the error term must lie within the bounds, the model is still estimated consistently with endogenous independent variables. To see this explicitly, say ξ_{jt} is a negative shock. In our model this will imply all ε_{ijt} will simultaneously receive negative shocks. However, as long as the shocks lie within the assumed bound δ, the partially identified set will still be consistent and contain the true demand distribution.

This feature remains true allowing for some proportion of the data to be contaminated, as we do in our setting. As with the bounded error models, so long as the proportion of assumed data contamination falls underneath our assumed upper bound for such decisions (we discuss how this can be determined systematically in the next section), the set of estimated preference distributions is still consistent and the true deterministic demand curve is contained within this set. Thus, our model will be robust to endogeneity concerns with respect to price or other features of the environment such as advertising or marketing.

2.3. Simulation

In order to illustrate our methodology, we study a simulated market where the firm or firms have information on consumer purchase behavior that they use to estimate a demand curve. The simulation gives us the ability to study the degree to which the partially identified demand output from the various proposed models links to underlying preferences. In our simulation we will have two products and an outside option and in each time period consumers decide which product to purchase (if any) given the specified price.

We simulate the preferences of 300 individuals from the population who obey the utility specification:

$$U_{ijt} = \alpha_{ij} + \beta_i p_{ijt}.$$

Here, α_i are product fixed effects which we use to aggregate preferences for all attributes except for price, as well as any other brand specific utility component. This is without loss of generality for our pricing problem since we assume that firms in this market do not change product attributes over time, except for price. Additional information about product attributes can only help refine the model further. Consumer i chooses product k at time t given the decision set d_{it} based on the decision rules in the four models just described.

We set the utility of the outside option for each person and time period to 0 and normalize the value of $\alpha_{i1} = 1 \ \forall i$ in order to obtain identification. Throughout the analysis there are two possible products, so there are two free preference parameters for each individual in the population. For this population, we draw α_{i2} from a uniform distribution on [0.5, 1.5] and β_i independently from a uniform distribution on the range $[-3.75, -1.75]$. We then simulate 208 time periods of choices (corresponding to four years of weekly data[23]) of decisions for each individual. In order to do so, we assume that both products are offered in every period and

---

[22] In the case that the true DGP is one where there is some probability p an observation is misclassified, it is possible that a particular individual has more than ϕ percent of observations misclassified. In that event, our estimate of δ (as described in Section 3) would increase to account for apparent larger deviations in stable preferences from misclassified observations not accounted for with ϕ. This again would lead to conservative estimates, as our partially identified parameters sets would reflect the larger δ values that would be necessary to rationalize these purchases.

[23] Most panel data sets available to researchers have 4 years of weekly purchase data. We have also experimented with fewer periods. If we have 50 weeks our overall conclusions do not change.
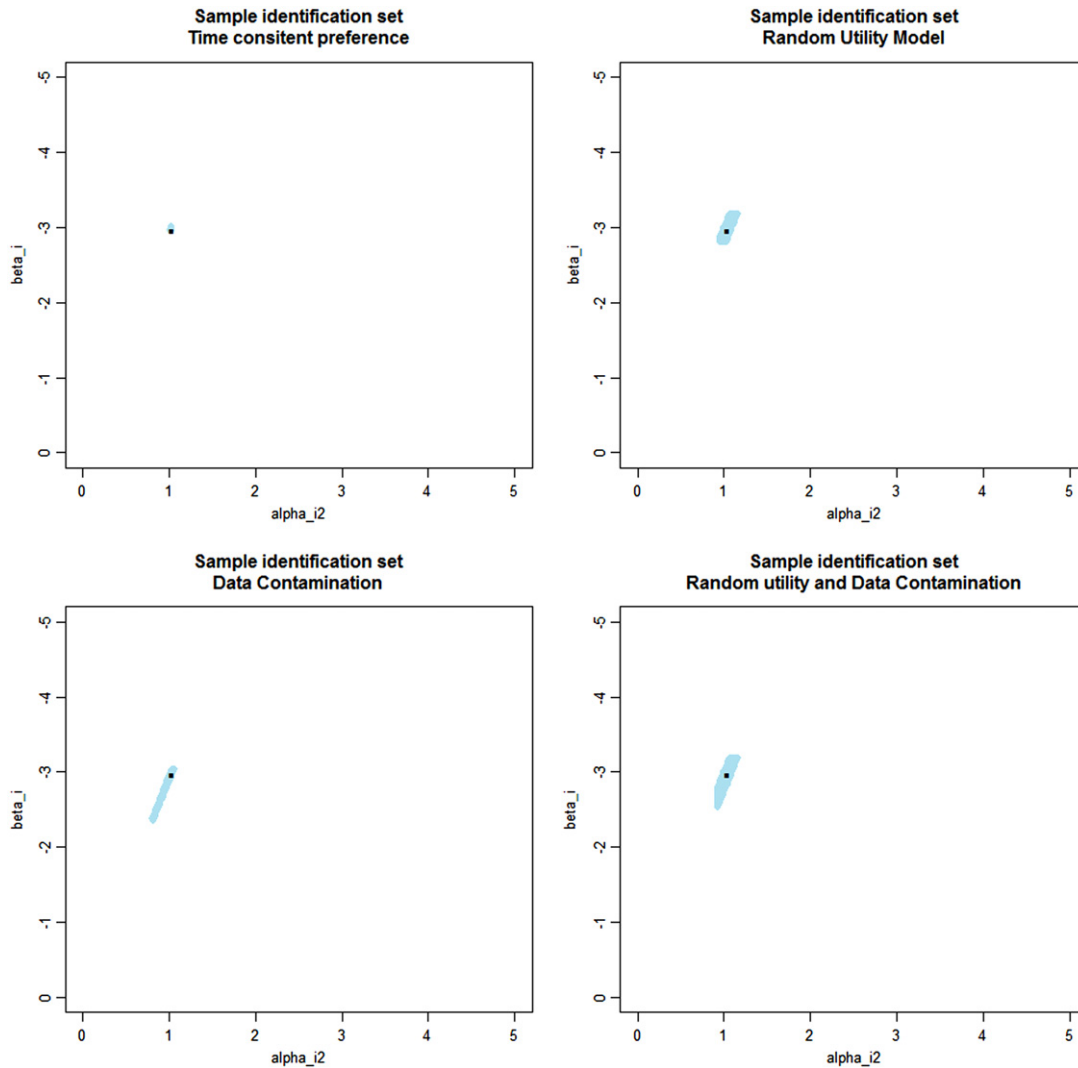
Fig. 1. Sample identification sets for a consumer. The black dot represents the true preference parameters, and the blue region represents the partially identified region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that their prices are drawn independently and uniformly from the range [0.1, 0.7]. We define the feasible identification region for (α_{i2}, β_i) to be $[0, 5] \times [-5, 0]$ so that the feasible identification region covers a large region of reasonable relative preferences.

We simulate data for each of the four choice models described above. For model 1 we simulate the data as just described. For the models with random utility shocks (models 2 and 4) we set $\delta = 0.10$. This implies that random utility shocks here are at most 10% of the base value of the preference for product 1. We allow there to be three types of consumers. The first 100 have i.i.d. errors. The second 100 have errors that are correlated across products. This correlation is generated in each period by first drawing ε_{1t} uniformly from the range $[-\delta, \delta]$, resulting in $\widehat{\varepsilon_{1t}}$. We then draw ε_{2t} uniformly from $[\widehat{\varepsilon_{1t}} - \delta, \delta]$ if $\widehat{\varepsilon_{1t}} > 0$ and from $[-\delta, \widehat{\varepsilon_{1t}} + \delta]$ if $\widehat{\varepsilon_{1t}} < 0$. The last 100 have errors that are correlated over time. This correlation is generated for each product by first drawing ε_{j1} uniformly from the range $[-\delta, \delta]$, resulting in $\widehat{\varepsilon_{j1}}$. We then draw ε_{j2} uniformly from $[\widehat{\varepsilon_{j1}} - \delta, \delta]$ if $\widehat{\varepsilon_{j1}} > 0$ and from $[-\delta, \widehat{\varepsilon_{j1}} + \delta]$ if $\widehat{\varepsilon_{j1}} < 0$. We then repeat for period three (and so on) replacing $\widehat{\varepsilon_{j1}}$ with $\widehat{\varepsilon_{j2}}$.[24] For models 3 and 4 we set the ϕ parameter to 0.10. This

implies that 10% of the data are contaminated. We also divide the consumers into three types: 100 have (up to) 10% of their decisions mis-recorded (randomly), 100 choose brand 1 independent of preferences (up to) 10% of the time and the other 90% of the time they make utility based decisions and the last 100 choose not to buy (up to) 10% of the time. However, these 10% are more likely (100 times) to occur after a purchase of brand 1 in time $t - 1$. These rules are intended to simulate data contamination.

Given each individual's choice data, we can partially identify her true parameters within this feasible set. The size of the partially identified set varies based on the observed choice behavior and the choice. Below, we give examples of partially identified sets for two consumers, based on the four choice models presented. The first individual (see Fig. 1) is an example of a consumer who purchases both products (and the outside option) at some point in time in our generated data for all four choice models. This allows us to refine the partially identified set of alternatives to a relatively small region for all four models. All the estimated identification regions contain the true parameters. The largest identification region will, by definition, occur for model 4 since the data generating process incorporates both data classification errors and utility shocks, implying less information about consumers' actual preferences than the other specifications. The second example, depicted in Fig. 2, is that of a consumer who never purchases product 2 in the

---

[24] The average (over consumers) correlation in the errors across brands is 0.544 and the average correlation in the errors between times t and $t + k$ is 0.483, 0.223, 0.108, 0.045 and 0.018 for $k = 1, 2, 3, 4$ and 5, respectively.
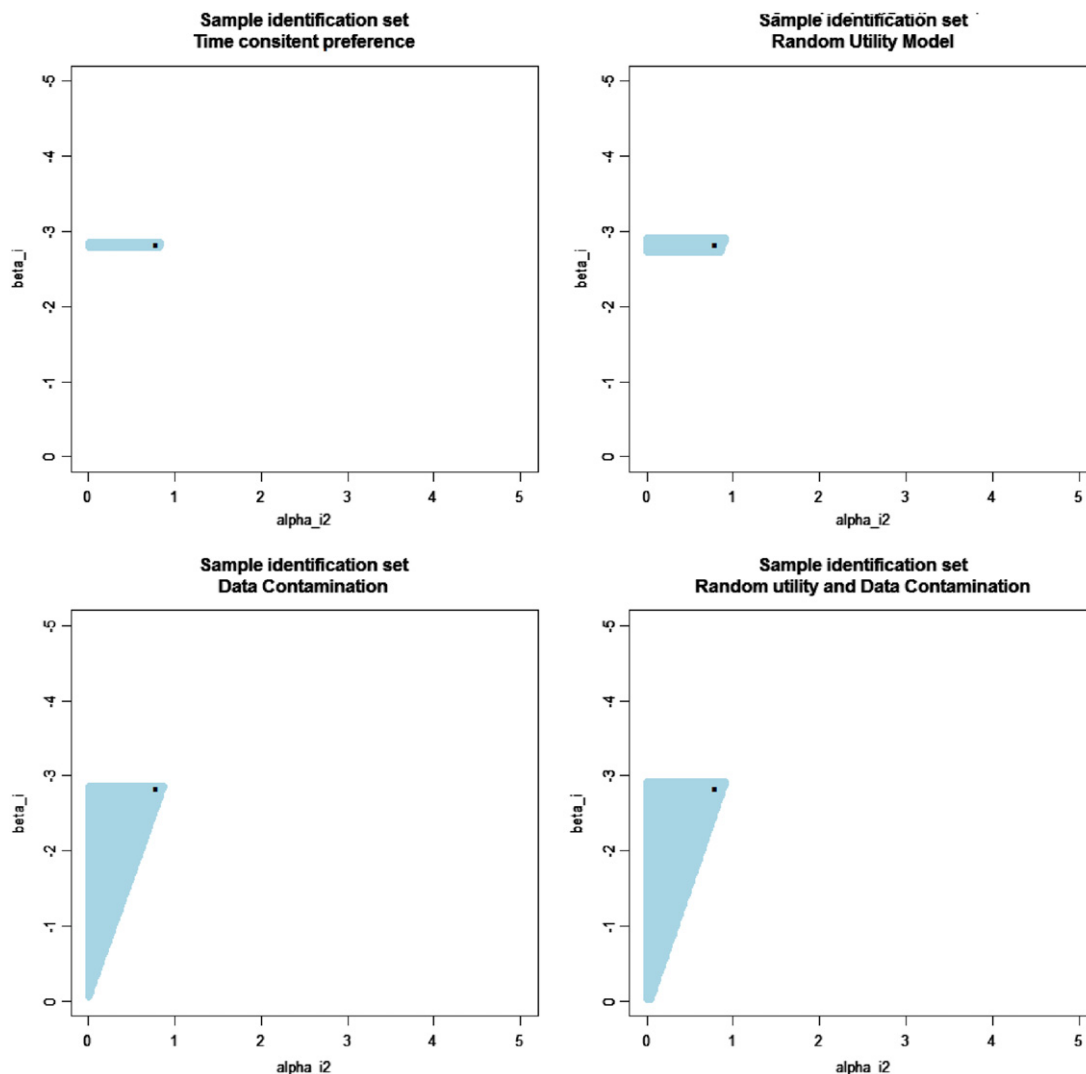
Fig. 2. Sample identification sets for a consumer. The black dot represents the true preference parameters, and the blue region represents the partially identified region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

entire data sample. Given these data we cannot achieve such tight bounds.

After we have found the partially identified set of preferences for all 300 consumers, we aggregate these regions in the manner described for each model to obtain a partially identified set for the joint distribution of tastes in the population. This distribution is generated in a two dimensional space (α_2, β space) and is used to estimate demand. The demand estimates are shown in two different formats. First, we consider a demand curve for each product if it were sold in isolation (without the other product) and then estimate demand with both products selling.

From Fig. 3 we obtain several insights. We correctly capture the true demand in the bounds for all of our models. The bounds are tightest in the model with the time-consistency assumption and are widest in model 4. Overall, the bounds are tight and can be informative for managerial decision-making. One way of thinking of this in terms of parametric discrete choice modeling is that any correctly specified model that describes these data must predict demand to be within the bounds specified.

Finally, in Fig. 4 we consider demand in a setting where both products are sold. These representations show that we do capture large parts of the demand curve with tight bounds across all models. Once again, the smallest bounds are with model 1 and the largest bounds are with model 4.

3. Estimation

The econometric framework just described assumes that the researcher and firm know the bound on the individual consumer-level preference shock δ as well as the maximum extent of data contamination ϕ. For cases where the firm and researcher know, or assume, δ and ϕ, there is a direct link from the data observed to the partially identified preference sets that are our primary econometric output, as described in Section 2. It is important to note that our model can be rejected by the data if the values of δ and ϕ imposed by the econometrician cannot explain the observed variation in the data.

In most empirical settings, researchers will have limited information about these parameters. In this section we propose a simple method to identify and estimate δ in a first stage that precedes the implementation of the framework set out in Section 2. To do this, we first resolve the issue of where the data contamination parameter ϕ comes from in an empirical setting, and then discuss the first stage estimation of δ conditional on that value of ϕ.

3.1. Data contamination

There are numerous reasons that data may be contaminated including, but not limited to, recording errors, non-response,
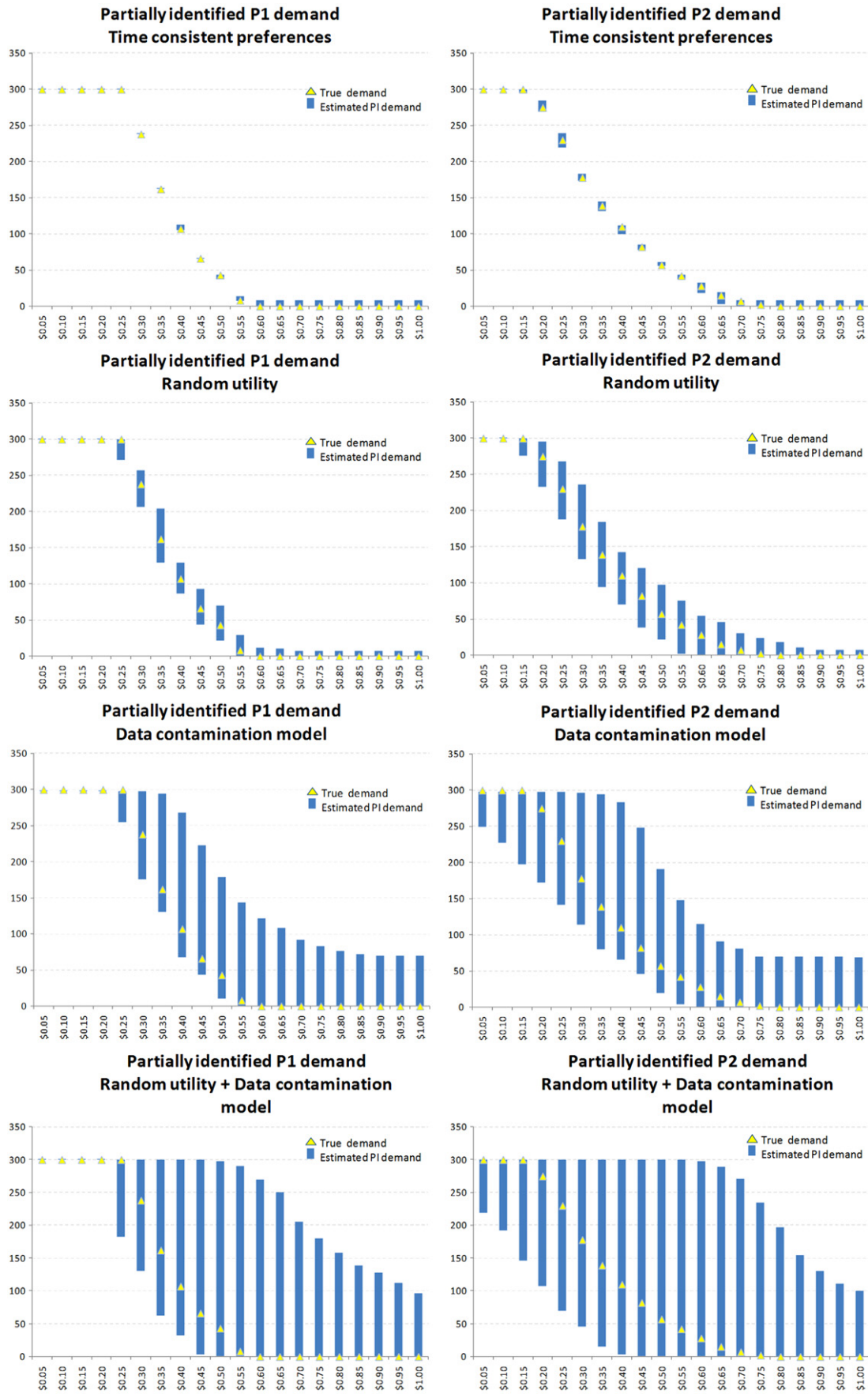
Fig. 3. Partially identified demand curves if products were sold in isolation. The rows represent the four models presented in this paper and the columns represent the demand curves for product 1 and product 2, respectively.

Fig. 4. Partially identified demand curves if both products are sold. The rows represent the four models presented in this paper and the columns represent the demand curves for product 1 and product 2, respectively.
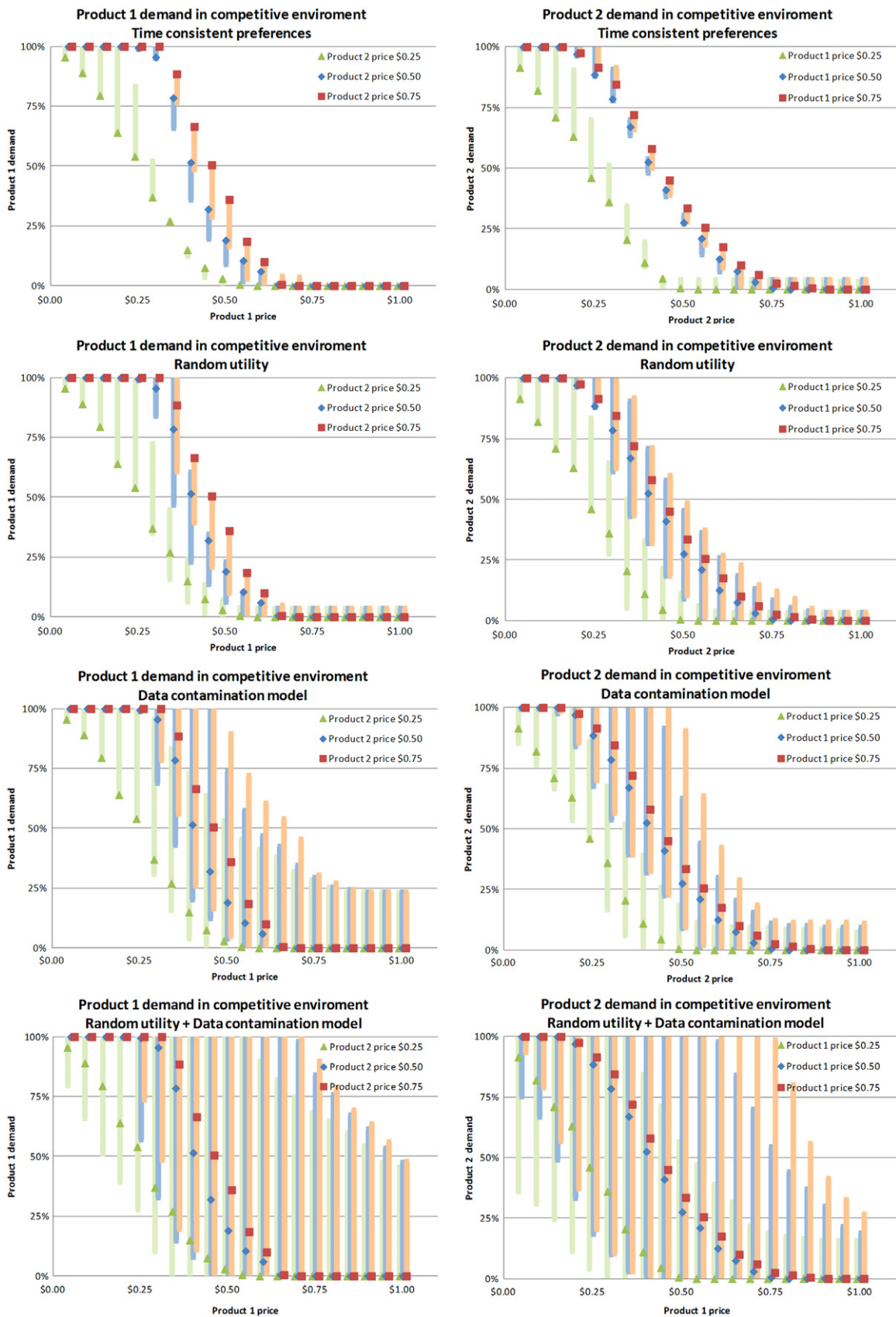
or interpolation/extrapolation. Numerous authors have explored the extent, and issues relating to identification and estimation in the presence of data contamination across myriad empirical settings (see for example Horowitz and Manski (1995), Keane (1997), Erdem et al. (1999) or Einav et al. (2010)). In our setup, we cannot separately identify both δ and ϕ using just the panel data because a large δ can be used to empirically justify the large departure from stable preferences represented by data contamination.[25]

In our approach, we lean on the ability to identify the extent of data contamination from past studies or a simple empirical investigation set up by the researcher or firm. We then use the panel data we observe to estimate δ conditional on ϕ.[26] Previous work, such as Erdem et al. (1999) or Einav et al. (2010) provide excellent sources for researchers to learn about the extent of data contamination. Additionally, in principle it is feasible in most empirical contexts for the firm or researcher to perform a similar type of validation study to get $\hat{\phi}$.

3.2. Estimating the error bound

Once we have a conjecture $\hat{\phi}$, we can estimate δ in a first stage that uses the same panel data used throughout the rest of the analysis. Define δ^* as the true value of δ we are trying to recover. We construct an estimator for δ^* that leverages both the time-series and cross-sectional variation across consumers in the panel data that we observe. As in Section 2, we illustrate the estimation with homogeneous δ^*. In practice we can condition estimation of δ^* on observable demographics without altering the methods described here.[27]

Define $\delta_{iT}^{\hat{\phi}}$ as the lowest δ that can rationalize consumer i's decisions when there are T time periods observed and $\hat{\phi}$ is the extent of data contamination. This is the minimum value of δ that results in the identified set for consumer i to be a non-empty set conditional on $\hat{\phi}$. Explicitly from the methodology in Section 2, this implies that the $\delta_{iT}^{\hat{\phi}}$ is the lowest value of δ, such that the set

$$\left\{ \theta : \theta \in \cap_{T_i^{\hat{\phi}}} \{ \theta : \theta \cdot x_{a_{it}^*} \geq \theta \cdot x_{a_{it}^{-*}} - 2\delta \ \forall a_t^{-*} \in d_{it} \} \right\}$$

is non-empty for some feasible $T_i^{\hat{\phi}}$, with the set of feasible $T^{\hat{\phi}}$ as defined in Section 2.[28] In addition to the assumption that δ is homogeneous conditional on observable demographics, we assume that the bound is tight for the potential purchase data we observe. That is, as we observe infinite data, the most extreme values of δ will be realized for each consumer. Formally, we have $\lim_{T \to \infty} P(\delta_{iT}^{\hat{\phi}} = \delta^*) = 1$.

Given these assumptions, we proceed as follows. In any sample we can define $\Delta_T^{\hat{\phi}} = \{\delta_{iT}^{\hat{\phi}}, i \in I\}$ and assume that the sample is ordered so that $\delta_{1T}^{\hat{\phi}} \leq \cdots \leq \delta_{IT}^{\hat{\phi}}$. A simple estimator of δ^* is $\hat{\delta} = max(\Delta_T^{\hat{\phi}})$. This will be biased downwards relative to δ^*. Though $\hat{\delta}$ would be consistent for δ^*, we follow the econometric literature on estimating boundaries and correct for this bias in our estimator (see the summary of current methods in Karunamuni and Alberts (2005)).

Denote this bias as γ. While the literature on estimating boundaries has a variety of sophisticated methods for estimating the edge of a distribution, Hall and Park (2002) discuss a simple m-out-of-n subsampling bootstrap method in which the estimate of the bias is $\hat{\gamma}_m^{HP} = \frac{\sum_i (\delta_{n-i+1} - \delta_{n-i}) K(i/m)}{\sum_i K(i/m)}$, where $K(\cdot)$ is a non-negative function, which can be interpreted as a kernel weighted average of the difference between consecutive values of δ. We adapt their approach, which is for continuous bounded random variables (as is generally the case in this literature) to our setting in which all admissible values of δ are from a discrete set (as we have a discrete set of types) in our model.[29] Despite this difference, we show in Monte Carlo exercises in Appendix B that our procedure effectively corrects for the bias in $\hat{\delta}$.

Define $\hat{f}(\delta_{iT}^{\hat{\phi}})$ as the empirical distribution of $\delta_{iT}^{\hat{\phi}}$ across I for fixed T, which will be a discrete distribution in our setting. Our estimator for γ is:

$$\hat{\gamma}_T = \sum_{\delta_{iT}^{\hat{\phi}} \in \Delta_T^{\hat{\phi}}} (\hat{\delta} - \delta_{iT}^{\hat{\phi}}) \hat{f}(\delta_{iT}^{\hat{\phi}}). \tag{13}$$

For this estimator to be consistent, we need $\lim_{T \to \infty} \hat{\gamma}_T = 0$. This is true as $\lim_{T \to \infty} f(\delta_{iT}^{\hat{\phi}}) = 0, \forall \delta_{iT}^{\hat{\phi}} \neq \hat{\delta}^*$ by our assumption of a common δ^* for the population in question. In other words both the simple estimator $\hat{\delta}$ and the bias-corrected estimator $\hat{\delta} + \hat{\gamma}$ are consistent for δ^*, but the latter is more conservative in the sense that it is less likely to underestimate δ^*.[30] The extent to which $\hat{\gamma}_T + \hat{\delta}_T$ is upward or downward biased in a given application depends on the interaction between the assumed form of the bias correction in Eq. (13) and the data. To evaluate the performance of this estimator we run a series of Monte Carlo simulations in Appendix B. Our simulations reveal that with more than 50 consumers and 100 time periods (reasonable values in the context of existing panel data sets) our estimator provides a reliable and conservative estimate for the true δ^*. Under a variety of data generating processes, the bias-corrected estimator has very few instances where estimated $\hat{\delta} < \delta^*$ and is usually quite close to δ^* coming from above. It is worth noting that, even when the bias correction we implement leads to an estimator that slightly overstates the true δ^* in the finite

---

[25] In our model as $\delta \to \infty$, all data can be rationalized as unbounded large "random" shocks and as $\phi \to 1$, all data can be rationalized as completely contaminated data.

[26] We note that, in principle, we could estimate the model by using an estimate of δ from "outside" the panel data we observe and then estimate ϕ with our data conditional on that value for δ. However, since it is easy to think of how one would construct data validation studies outside of the panel data, but difficult to think about studies that would inform the extent of preference shocks, we believe that the approach outlined in this section is more practical for most empirical settings.

[27] It is important to point out that as T becomes large and we observe more purchase data per consumer, it becomes more attractive to try and estimate unobserved heterogeneity in δ^*, conditional on a demographic profile. In the limit as $T \to \infty$ (and there is enough price variation), we identify the true δ^* for each individual. The estimation here recognizes that we do not observe infinite data in reality, and uses homogeneity conditional on demographics to help identify δ^* in the relevant group of interest.

[28] If a set of preferences is feasible any given $T_i^{\hat{\phi}}$ then it is feasible in general.

[29] The literature on boundary estimation considers a setup where the econometrician observes N draws from a continuous univariate distribution F with a unknown and finite upper boundary. Our estimator below is a discrete analog to these methods, since the distribution of $\hat{\delta}_{iT}$ is discrete. As in Siddiqui (1960) and Bloch and Gastwirth (2002) we take the difference between the mth highest draw (of the N draws) from F and the highest draw, and weight by $\frac{1}{M}$. As in the Hall and Park (2002) estimator discussed in the text, we consider a kernel weighted average of different values of m ranging from 2 to M out of the N total draws. This allows us to use the range of draws while not having the worry about the fact that, in discrete data, the difference between the j and $j + 1$ highest draws could often be zero, as one would have to if using Hall and Park (2002) directly.

[30] Estimating the choice model defined in this paper with a value of δ less than the true δ^* could result in biased estimates of individual types (i.e. partially identified sets that do not contain true preferences). On the other hand, estimating the choice model with a value of δ greater than the true δ^* will lead to a loss of efficiency (the partially identified sets will be larger), but not a bias. This is one reason that we feel it is important to correct for the bias in $\hat{\delta}$ with $\hat{\gamma}$.

sample analysis, in our framework this leads to more conservative estimates in the form of larger partially identified preference sets.

It is important to note that, with the assumptions maintained in this section, δ^* for the population (conditioning on any observable demographics) is identified by the purchases of a given consumer i as $T \to \infty$ and there is sufficient variation in prices. With finite data on T, a larger number of consumers I improves the precision of $\hat{\delta}$ and $\hat{\gamma}$. Also, we note that while the model can be rejected by the data for the general case where ϕ and δ are imposed by the econometrician, for the estimation methodology proposed here the model will not be rejected by the data since δ is estimated as the minimum rationalizable value.

4. The firm problem

We now demonstrate how a firm which has partially identified demand in the manner described thus far can make strategic decisions in the face of the resulting ambiguity about consumer preferences. We focus on perhaps the quintessential firm problem: how to set prices. We examine the cases of a monopoly and a duopoly.

4.1. Monopoly

A monopolist observes a panel of individual decisions and makes a decision on what price to charge to the same target population that composes the sample. Given the framework above, the monopolist observes a range of feasible type distributions that characterize the population and uses that information to arrive at a pricing decision. The monopolist wishes to maximize profits, but cannot necessarily do so in the traditional way because he does not know the distribution of types in the population and thus can only partially identify demand for each potential price.

Given that the firm does not know the distribution of types, expected profit maximization is not possible. Therefore, we must take a stand on how the firm makes its pricing decision. As in Bergemann and Schlag (2007), we examine the monopolist's problem using the minimax-regret criterion, which does not incorporate subjective beliefs on the state space by the decision-maker. Instead, this decision-making criterion is to minimize the largest possible "distance" from what the actual best choice would have been, were it to know the true state of the world *ex post*. It is conservative in the sense that it analyses the maximum regret (that is, the maximum distance from the ideal value over all possible states), but less conservative than a pure maxmin criterion. This is because the minimax-regret criterion accounts for deviations from possibly very good outcomes as well as just considering the worst case scenario (as would the maxmin criterion). We assume that the firm solves a constrained minimax-regret problem where the vector of possible prices chosen is fixed, instead of allowing for random pricing or menu pricing.

The monopolist in our setting has data on past purchase decisions by the population of consumers and seeks to maximize profits in a counterfactual setting in which prices can be set at levels not yet observed in the data for a given set of products. In our setting, the fundamental state is the distribution $F(\theta)$ describing the population of discrete types. If the monopolist knew this state exactly, it could easily construct a demand function for its product. We will denote demand as $D(F(\theta), p)$. Here, p is a price vector and $D(F(\theta), p)$ is the demand vector, where both quantities are vectors because we assume that the monopolist can sell multiple brands. A monopolist's regret is a mapping from any chosen price and given distribution of preferences into a scalar which measures how "far" the profits resulting from the chosen price are from the profits that would result from the optimal price if the candidate distribution of preferences were the true distribution. Maximum regret for the monopolist, given a choice of a price vector p and the potential

distributions of types $H[F(\theta)]$, is defined as:

$$R(p, H[F(\theta)]) = \max_{F(\theta) \in H[F(\theta)]} p^*(F(\theta))D(F(\theta), p^*) - pD(F(\theta), p).$$

(14)

The first term in Eq. (14) denotes the optimal profits for the monopolist if it knew that the true distribution of types was $F(\theta)$. Here, p^* is the price vector that implements this ideal profit level. From now on, we will denote the ideal profit level given a state $F(\theta)$ as $\pi^*(F(\theta))$. In addition, we will simplify notation by alluding to the quantity $pD(F(\theta), p)$ as $\pi(p, F(\theta))$, the profits earned by the monopolist in state $F(\theta)$ given some chosen price vector p. The key empirical challenge is estimating $\pi(p, F(\theta))$ for every possible distribution of types.

For every potential price, we calculate demand for each distribution of preferences. We find the optimal price as the price with the highest demand for any given distribution and regret for any other price is the difference between the profit under that price and the optimal price. Once we have calculated the regret for every potential price and distribution combination, we can easily calculate the maximum regret for any price (maximum over distributions). Then, we choose the price which minimizes the maximum regret. To be clear, we can define the monopolist's minimax-regret, given the identification region $H[F(\theta)]$, as:

$$MMR(H[F(\theta)]) = \min_p \max_{F(\theta) \in H[F(\theta)]} \pi^*(F(\theta)) - \pi(p, F(\theta)). \quad (15)$$

For any combination of $(p, F(\theta))$, a monopolist's regret will stem from either overpricing or underpricing based on whether p is greater than or less than p^*, respectively. In the Bayesian setup, this overpricing and underpricing for each $(p, F(\theta))$ pair is weighted by a subjective Bayesian prior over $H[F(\cdot)]$ and regret minimization with respect to this weighting is equivalent to expected profit maximization.

We denote the minimax-regret solution as p^{MMR}. Since the minimax-regret state space is directly defined by the econometric exercise, the state space is complex in the sense that it is impossible to obtain an analytical solution to this problem. This is an important way in which the current paper differs from that of Bergemann and Schlag (2007), since their model relies on the set notion of an ε neighborhood, given some size ε, and finds solutions analytically. In practice, implementing the minimax-regret solution requires a multi-stage algorithm given $H[F(\theta)]$. First, for each distribution $F(\theta)$ and each feasible price vector p, we compute the demand vector $D(F(\theta), p)$. Then, we compute the ideal profit given for each $F(\theta)$. Afterward, we compute the maximum regret for each price vector over the identification region for feasible distributions. Finally, we minimize these maximum regrets over all possible price vectors.

Before we move on to the oligopoly problem, we present a stylized example meant to illustrate the way one can think about the monopolist's minimax-regret problem.

4.1.1. Stylized example

This section presents a simple example of the monopolist minimax-regret problem for one good. We assume that there is one preference parameter which translates directly into demand with feasible values in [1, 2] given our econometric input. This is a much simplified version of our model, in which we must first translate partially identified distributions of types into demand in a non-trivial way. The benefit is that it provides intuition for minimax-regret in a very simple framework. Suppose, that given the possible distributions in $H[F(\theta)]$, the monopolist knows that, for a given p, the range of demand is $[1 - 2p, 1 - p]$, where this range comes from the mapping $D(F(\theta), p)$, taken over $H[F(\theta)]$ for

each p. We will assume the marginal cost equals zero for simplicity. The monopolist's minimax-regret problem is:

$$\min_p \max_{F(\theta) \in H[F(\theta)]} \pi^*(F(\theta)) - \pi(p, F(\theta))$$

$$\iff \min_p \max_{\beta \in [1,2]} \frac{1}{4\beta} - p + \beta p^2.$$

Now, when the monopolist solves for his maximum regret over β given his choice of price, he only has to consider two states, $\beta \in \{1, 2\}$. This is because, given the optimal price $p^*(F(\theta))$, the profit function is monotonically decreasing on $\mathbb{R}^+$ going in both directions from that optimum since profit is a quadratic function of price. This implies that for any given price, the maximum regret will be one of the endpoints of the range of β, since the optimal price given β is monotonically decreasing in β. Thus, we can map the range of β directly into a range of optimal prices, $p^*(\beta)$, and for any given p, maximum regret will occur at the maximum possible distance from a feasible $p^*(\beta)$, which will always correspond to an extreme value of β. In our example, the regret functions for p given $\beta \in \{1, 2\}$ are:

$$R(\beta = 1, p) = \frac{1}{4} - p + p^2$$

$$R(\beta = 2, p) = \frac{1}{8} - p + 2p^2.$$

The first function is minimized with zero regret at $p = \frac{1}{2}$, while the second is minimized at $p = \frac{1}{4}$. Each function is monotonically increasing in both directions from its respective minimum, so we know that the minimax-regret must occur in the range $[\frac{1}{4}, \frac{1}{2}]$ at the point where both of these functions have identical regret values given p. This occurs when:

$$\frac{1}{4} - p + p^2 = \frac{1}{8} - p + 2p^2 \Rightarrow p^{MMR} = \frac{1}{\sqrt{8}}.$$

The solution is easy to verify. If p is increased or decreased from p^{MMR}, the maximum regret increases because one of the two regret functions corresponding to $\beta \in \{1, 2\}$ must increase. This provides some insight into the *MMR* solution that we will derive in our pricing experiments in the next section. For a given set of distributions $H[F(\theta)]$, there will be an extreme distribution that corresponds to the minimum and maximum demand for a given price. For that price, it will then be possible to compute the maximum regret, which will then be compared over all prices to derive the final solution, which will balance the potential losses from pricing low in a low elasticity state and pricing high in a high elasticity state.

4.2. Oligopoly

In addition to the monopolist's problem specified above, we analyze a static oligopoly game. In this setting, every firm shares the same information set and evaluates payoffs according to minimax-regret over $H[F(\theta)]$, given the other firms' prices p_-. This extends the assumption that the firm and the researcher have the same information set to one where both firms and the researcher have the same information set, which we believe is more reasonable in situations where firms observe similar limited data to base their pricing decisions on.[31] A further more detailed model where firms have only partial information about the other firms' information sets (or their perceptions of the distribution of types) would be interesting, but for now we stick to this base case and leave this extension to future work.

Let firms be indexed by j corresponding to J different sets of brands. The firms play a game where each evaluates outcomes by minimizing maximum regret over possible price vectors given

their opponents' prices. The firms evaluate maximum regret for p_j given the opponents' price vector p_{-j} as follows:

$$R(p_j, p_{-j}, F(\theta)) = \max_{F(\theta) \in H[F(\theta)]} p_j^*(F(\theta), p_{-j}) D(F(\theta), p_j^*, p_{-j})$$

$$- p_j D(F(\theta), p_j, p_{-j}). \quad (16)$$

Here, the firm evaluates regret at a given state of nature conditional on his opponents' prices. His minimax-regret given p_{-j} is:

$$MMR(H[F(\theta)], p_{-j}) = \min_{p_j} \max_{F(\theta) \in H[F(\theta)]} \pi^*(F(\theta), p_{-j})$$

$$- \pi(p_j, p_{-j}, F(\theta)) \quad (17)$$

$\pi^*(F(\theta), p_{-j})$ is the ideal profit for firm j, given a specific distribution of types drawn from the identification set and p_{-j}. $\pi(p_j, p_{-j}, F(\theta))$ is the profit for firm j given the type distribution and opponent's price. In the game that the firms play, the action space is the set of feasible non-negative prices, and we restrict each firm to the use of pure strategies. The game is one of complete information between players in the sense that each firm knows the uncertainty faced by the other with respect to the distribution of types. We assume that both firms have common knowledge and look for a pure strategy Nash equilibrium in price vectors. We say that firm prices p^{NE} are a Nash equilibrium if the following best response conditions are simultaneously satisfied:

$$p_j^{NE} \in \arg\min_{p_j} \max_{F(\theta) \in H[F(\theta)]} \pi^*(F(\theta), p_{-j}^{NE}) - \pi(p_j, p_{-j}^{NE}, F(\theta)),$$

$$\forall j \in J. \quad (18)$$

In our simulation described in the next section we find a pure strategy equilibrium using the best response curves of each firm to his opponent's price, given the identification region $H[F(\theta)]$.

5. Empirical analysis

The purposes of this section are (i) to compare the results of our model to those of the most commonly used discrete choice models and (ii) to show an example of how our method can be applied to data. We will start with a simulation experiment in Section 5.1 where we show that when the underlying data violate the i.i.d. assumption of the standard discrete choice models, our model provides a more reliable and robust pricing recommendation.

We then consider panel data from milk purchases from two competing retailers in Section 5.2. We show that our methodology is applicable in this real-world setting and that it returns sensible counterfactual recommendations.

5.1. Simulation experiment of the firm problem

In order to illustrate our methodology, we study a simulated market where the firm or firms have information on consumer purchase behavior that they use to determine how prices should be set. The simulation gives us the ability to study how solving the firm problem with our method compares to what the firm would do if it knew the true distribution of preferences in the population. In addition, it allows us to study the predictions of our model compared to more familiar models, such as a mixed logit model, when estimated with the same data.

We simulate the preferences of 100 individuals with utility $u_{ijt} = \alpha_{ij} + \beta_i p_{ijt}$, as was done in Section 2.3.[32] We assume that consumer i chooses product k at time t given the decision set d_{it}

---

[31] Our empirical example is one situation where this may be reasonable.

[32] For clarity, as was done in Section 2.3, we set the utility of the outside option for each person and time period to 0 (location invariance) and normalize the value of $\alpha_{i1} = 1 \ \forall i$ (scale invariance) for identification. Throughout the analysis

Table 1
Results of monopoly pricing from simulation. See text for details on simulation design.

Consumer type	Pricing model	Price 1	Price 2	% of optimal profits
i.i.d. shocks				
	ex post efficient	0.27	0.36	
	Mixed logit	0.27	0.45	90
	Minimax-regret	0.27	0.36	99
Brand correlated shocks				
	ex post efficient	0.26	0.36	
	Mixed logit	0.36	0.36	84
	Minimax-regret	0.27	0.36	98
Time correlated shocks				
	ex post efficient	0.28	0.39	
	Mixed logit	0.90	0.90	0
	Minimax-regret	0.27	0.36	96

based on the decision rules in model 2 described in Section 2. We set $\delta^* = 0.20$ (for this exercise, $\phi = 0$, i.e. there is no data misclassification). Given our parameterization, this implies that random utility shocks here are at most 20% of the base value of the preference for product 1. We simulate three types of consumers. The first type have i.i.d. errors drawn from a uniform distribution between $[-\delta^*, \delta^*]$. The second type have errors that are correlated across products. This correlation is generated in each period by first drawing ε_{1t} uniformly from the range $[-\delta, \delta]$, resulting in $\widehat{\varepsilon_{1t}}$. We then draw ε_{2t} uniformly from $[\widehat{\varepsilon_{1t}} - \delta, \delta]$ if $\widehat{\varepsilon_{1t}} > 0$ and from $[-\delta, \widehat{\varepsilon_{1t}} + \delta]$ if $\widehat{\varepsilon_{1t}} < 0$. The third type have errors that are correlated over time. This correlation is generated for each product by first drawing ε_{j1} uniformly from the range $[-\delta, \delta]$, resulting in $\widehat{\varepsilon_{j1}}$. We then draw ε_{j1} uniformly from $[\widehat{\varepsilon_{j1}} - \delta, \delta]$ if $\widehat{\varepsilon_{j1}} > 0$ and from $[-\delta, \widehat{\varepsilon_{j1}} + \delta]$ if $\widehat{\varepsilon_{j1}} < 0$. We then repeat for period three (and so on) replacing $\widehat{\varepsilon_{j1}}$ with $\widehat{\varepsilon_{j2}}$.

We proceed first by partially identifying each consumer's preferences as was illustrated in Section 2.3. Given each individual's choice data, we can partially identify true parameters as being within this feasible set. Next we turn to the industry pricing problem. To evaluate our model we compare it with two benchmarks: (1) *ex post* efficient prices based on true parameters and (2) a mixed logit with multivariate normal mixing.[33]

Monopoly

We begin by considering a multi-product monopolist possessing the purchasing decisions of each set of 100 consumers over 100 periods. The firm must now set prices for each of its goods. In Table 1 we show three options for the optimal prices: (1) *ex post* efficient prices, from the simulated values of each individual; (2) optimal prices from the mixed logit model where we consider the prices that maximize expected profits; and (3) optimal prices from the minimax-regret model as described in Section 4.

---

there are two possible products, so there are two free preference parameters for each individual in the population. For this population, we draw α_{i2} from a uniform distribution on $[0.5, 1.5]$ and β_i independently from a uniform distribution on the range $[-3.75, -1.75]$. We then simulate 100 time periods of choices (corresponding to about two years of weekly data) of decisions for each individual. In order to do so, we assume that both products are offered in every period and that their prices are drawn independently and uniformly from the range $[0.1, 1.0]$. We define the feasible identification region for (α_{i2}, β_i) to be $[0, 5] \times [-5, 0]$ so that the feasible identification region covers a large region of reasonable relative preferences.

[33] Our mixed logit model is specified as $u_{i,j,t} = \alpha_{i,j} + \beta_i P_{j,t} + \varepsilon_{i,j,t}$ for $j = 1, 2$ and $u_{i,0,t} = \varepsilon_{i,0,t}$. We assume the ε are distributed i.i.d. Type 1 Extreme Value. For heterogeneity we assume $(\alpha_{i,1}, \alpha_{i,2}, \beta_i) \sim N((\alpha_1, \alpha_2, \beta), \Sigma)$. We use simulated maximum likelihood with 100 draws to estimate $\alpha_1, \alpha_2, \beta$ and the Cholesky decomposition of Σ (we follow the estimation procedure in Revelt and Train (1998)).
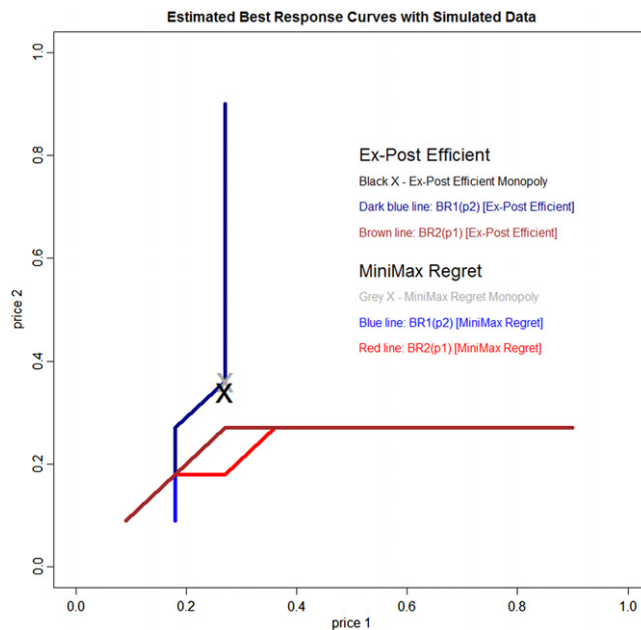
Fig. 5. Best response curves in duopoly simulation. See text for details on simulation design.

Note that for the mixed logit and the minimax-regret model we consider prices in 0.09 increments between 0.09 and 0.90.[34]

The first set of results in Table 1 are for the case where consumers are drawn with i.i.d. errors. In this case the minimax-regret model estimates prices that are close to *ex post*, with firms earning nearly 100% of potential *ex post* profits. The mixed logit optimal prices are also close to the *ex post* efficient ones, and recover nearly 90% of the optimal *ex post* profits. This suggests that both models recommend nearly *ex post* efficient prices when consumers have i.i.d. error draws. The second and third sets of results in Table 1, where consumers have either brand correlated error shocks or time correlated error shocks, are noticeably different. Here while the minimax-regret model sill recommends prices close to *ex post* efficient prices, the optimal prices from the mixed logit are far too high. This is particularly evident for consumer type 3 with time correlated shocks.

Duopoly

We now turn to the case of multiple single product firms in a differentiated goods industry. We focus on the duopoly case, pool all 300 simulated consumers from the monopoly experiment (100 consumers of each type) and now assume that the two products are sold by two different firms. Each firm must now choose the price for its good taking into account what the other firm will do. We solve for the equilibrium of the pricing game by finding the intersection of the firms' best response curves, depicted in Fig. 5. We also show the true *ex post* efficient response curves for the firms in this figure.

Here the minimax-regret model estimates best response curves close to the *ex post* efficient best response curve. The model recommends duopoly prices of 0.18 for each product. In comparison, the mixed logit model here would recommend duopoly prices of 0.90 for each product. This occurs because, as before, consumers in this simulation have non-i.i.d. error draws.[35]

---

[34] Using a finer grid did not materially affect the qualitative findings in this section.

[35] As a point of further comparison, we have estimated the optimal prices with a different non-prior based decision rule: maxmin. Under maxmin preferences, the
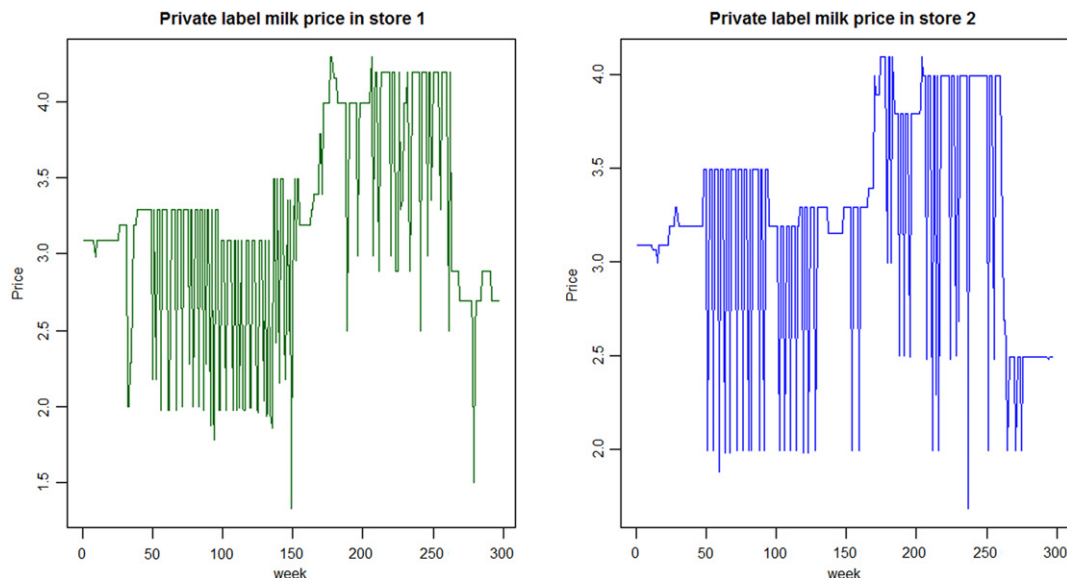
Fig. 6. Observed prices for private label, 1 gallon milk for two competing stores in Pittsfield, MA.

Table 2
Simulation results for monopoly and duopoly pricing when δ is estimated. Actual δ used to generate data is 0.20. See text for more details on simulation design.

δ	Monopoly		Duopoly	
	Price 1	Price 2	Price 1	Price 2
Assumed to be 0.2	0.27	0.36	0.18	0.18
Estimated to be 0.285	0.27	0.36	0.18	0.18
Assumed to be 0.5	0.36	0.45	0.27	0.27

Simulation results with estimating δ

The simulation results for the minimax-regret model for both the monopoly and duopoly results assume that we know the true value of δ. In this section estimate a value of δ, as described in Section 3.2, and then determine optimal prices from the minimax-regret model. We use the same 300 consumers from above and re-estimate the monopoly and duopoly prices with an estimated value of δ. Recall that the true value of δ used to generate the data was 0.20. For these data, the lowest value of δ that can rationalize all observed decisions is 0.19. With the bias adjustment, the estimated value of δ is 0.285, which is a conservative estimate of 0.20. We report the recommended prices when we assume we know the true δ and when we estimate it to be 0.285 in Table 2. For both the monopoly and duopoly cases, the suggested prices based on the estimated δ are the same as when we know δ's true value. This need not be the case. For example, if we assume $\delta = 0.5$ and again make optimal recommended prices, they will differ substantially as the final row in the table illustrates.

5.2. Field data analysis

The purpose of this section is to provide an illustrative example to show how our method can be used to study pricing in an empirical context with field data. We apply our model to six years of IRI panel data from the fluid milk category in Pittsfield, MA. Milk is a frequently purchased, non-storable, product category in which the top selling UPCs are private label brands. One reason for considering milk is that it is non-storable and is thus unlikely to be stockpiled. This is important because stockpiling, in addition to creating complex error structures that could invalidate standard logit model assumptions, will generate a dynamic choice process, the modeling of which is beyond the scope of this paper (for papers that do model this behavior see Erdem et al. (2003) or Hendel and Nevo (2006)).

We choose to study the pricing decisions of two neighboring retailers that have the highest unit sales in the IRI panel data in Pittsfield, MA (see Bronnenberg et al. (2008) for a detailed discussion of these data). As is common for retail milk prices, both stores charge the same price for all private label, one gallon milk products independent of fat content (Khan et al. (2012)). For these two stores we observe nearly 6 years (297 weeks) of panel data for 396 panelists.[36] Fig. 6 displays the average weekly prices for each store over the sample. The median prices in the data for the two stores over the 6 years are $3.20 and $3.22 per gallon, respectively. The correlation in prices across stores is 0.44. On average, 9% of panelists buy from store 1 and 17% of panelists buy from store 2 in a given week. We find 55% of the panelists make at least one purchase in each store over the six years.

To compute optimal prices we must have a measure of store-level marginal costs. The majority of this cost is likely to be the wholesale price the stores pay for milk. As a proxy for wholesale prices, we collect average monthly Co-op prices[37] in Massachusetts for 2008 (data from USDA's Agricultural Marketing Services). Note this time period overlaps with the last 6 months of our panel data. In the overlapping time period we find the average wholesale price is $0.60 per gallon and the average retail prices at the two stores are $2.68 and $2.45 per gallon, respectively.

---

firm chooses the price that maximizes its profits given the realization of worst-case demand for that price, selected from the set of feasible demand curves (see e.g. Gilboa and Schmeidler (1989)). We find that, consistent with the arguments discussed in Bergemann and Schlag (2008) and Manski (2005), the maxmin criterion tends to prescribe overly conservative decisions. For example, in the duopoly case we find that the maxmin duopoly prices are 0.09 and 0.09 and the model recovers only about 50% of the ex post optimal profits.

[36] We consider panelists who make at least 50 milk purchases from this store in six years.

[37] We will assume that this is each firms' marginal costs. It is realistic to assume that each pays the same wholesale price, but unrealistic to assume that this wholesale price is the entirety of each firm's marginal cost, and therefore we view our optimal prices likely as being lower bounds on the true optimal prices for each store.

To apply our model in this setting we first need to calibrate a value of ϕ. Einav et al. (2010) study a panel scanner data setting similar to our own, and report two estimates for the level of data contamination. First, they report that about 20% of purchase trips in the panel scanner data that they observe are incorrectly recorded. Second, they report that approximately 50% of actual purchase trips are omitted in the scanner data. In our data, consumers purchase from either store in 26.5% of weeks. From this analysis, we construct two alternative measures of the extent of data contamination that might be present in our data, ϕ, defined as the maximum proportion of purchase recording errors per individual. Our first measure assumes that only the first type of error (observed purchases that are incorrect) occurs. This suggests that 20% of the 26.5% of purchases made every week are contaminated, or that 5.3% of all purchase/no-purchase decisions are contaminated. Our second measure allows for data contamination stemming from purchases that are not recorded, suggesting that 10.6% (50% of 26.5% − 5.3%) purchases were never recorded in the scanner data in our context. As there is some room for interpretation in the way the Einav et al. (2010) results link to our setting, we estimate δ and study firm pricing in the two alternative cases where $\phi = 0.053$ and $\phi = 0.106$.[38]

While the Einav et al. (2010) study is the best source for calibrating a value of ϕ in our context, it is not ideal. This is because they focus on Nielsen Homescan data in which consumers record their purchases at home. Although some panelists in the IRI data we use also record their purchases at home, many scan them in the store (see SymphonyIRI (2012) or Bronnenberg et al. (2008) for more information on the IRI method of data collection).[39] It is likely that the extent of data recording error is greater for those scanning at home as opposed to at the store.[40] Therefore, for internal consistency and robustness we also investigate firm pricing for smaller ϕ equal to 0.026, or half the lowest value calibrated from the Nielsen Homescan data.[41]

For $\phi = 0.053$, we find that the lowest value of δ that can rationalize all observed decisions in the data is 0.37. Using the bias correction methodology discussed in Section 3.2, the estimated value of δ as 0.442. Using $\delta = 0.442$ and $\phi = 0.053$ we estimate preferences and compute the minimax-regret best response curves for each store (assuming a wholesale price of $0.60 per gallon). Our estimates yield a unique pure strategy equilibrium where one store charges $2.40/gallon and the other charges $2.35/gallon for private label milk.[42] We believe that these are reasonable estimates for this market since they are similar to the observed price levels in the raw data, where the prices between the stores are often close together (the median weekly price difference between the two stores is zero). Repeating the analysis for $\phi = 0.106$ yields

equilibrium price estimates of $2.50/gallon and $2.35/gallon for stores one and two respectively. For the lower value of ϕ, 0.026, estimated equilibrium prices are $2.40/gallon and $2.30/gallon. This suggests that our price estimates are relatively robust to the possibility that the IRI data that we perform the analysis with has a lower amount of data recording errors than the Nielsen Homescan data that we use to calibrate ϕ.

6. Conclusion

This paper presents an econometric framework that partially identifies consumer preferences and market demand under weak assumptions in a setting with panel data. The identification restrictions we make combine non-parametric methods from the Samuelsonian revealed preference tradition with an attribute based product representation commonly used in the discrete choice literature. Overall, the cost of maintaining very weak assumptions on the structure of consumer utilities is that we produce bounds on the distribution of preferences instead of a point estimate, as almost all work in this area does. However, the predictions that we can make with our methodology are more credible than those made under the traditionally strong assumptions found in this literature. We view this work as complementary to past work since the results from an approach with more (correct) assumptions should fall within the bounds that our model provides, while the placement of point identified results relative to our bounds shed light on the potential direction of bias in these results.

The purpose of developing and analyzing our econometric framework is to better model and understand how firms price in an environment with limited information leading to a high level of uncertainty. We characterize this uncertainty here with the notion of ambiguity though this is just one possibility for how firms with limited information could price (relying more on simple heuristics is another possibility). We model the firm pricing problem under ambiguity, borrowing from the theory literature, and link this to the econometric framework by assuming that the ambiguity the firm faces is described by the partially identified distribution of the demand curve. We investigate this joint framework in both monopoly and oligopoly settings to increase the scope of the framework.

Perhaps the most substantial contribution of this paper is to develop a joint theoretical and empirical framework that is a credible alternative to the full information mixed logit and expected profit maximization workhorse model used to analyze firm decision-making. Through simulations we show how our framework performs relative to this standard approach, concluding that a robust pricing framework can perform as well or better in terms of predicting what prices firms actually choose. We also illustrate how the methodology can be applied in an actual empirical setting. In situations where firms have limited information and the combined logit-expected profit maximization approach seems implausible, we provide a credible framework that can be used to study standard industrial organization problems without assuming such a high bar for firm knowledge and decision-making.

---

[38] If we assume that the probability a given observation is misclassified equals 0.053, as opposed to the upper bound on the proportion of misclassified observations per individual, a binomial calculation reveals that, with 100 purchases per individual in the panel, 99.3% of individuals will have 11 misclassified observations or less, so only very few individuals would violate the higher bound $\phi = 0.106$. The firm pricing results are quite robust to doubling ϕ to 0.106, as shown below.

[39] As discussed in Bronnenberg et al. (2008), of the panelists in our data, 47% only scan purchases in the store and 53% scan their purchases both in the store and at home.

[40] Additionally, the type of data recording error may differ across the two data collection methods as well. Since only the extent of data recording error matters for our model, we ignore this additional difference in data recording errors in the current paper.

[41] In 2009, IRI and Nielsen merged their panel data collection operations and, from that point on, these data sets (i) reflect the same panel of consumers and (ii) come exclusively from home scanned data (see Nielsen (2009)).

[42] The bootstrap standard errors, based on 5000 iterations, for the prices are 0.02 and 0.05, respectively.

Appendix A. Cross-sectional identification: for online publication

In many cases a panel sample may not be representative of the entire market. In this appendix we present a framework to use cross-sectional identification in conjunction with panel identification to form bounds on demand parameters when the panel is not representative of the population. We add two elements to the models outlined in Section 2. First, we posit that in addition to observing the panel across each of T time periods, we now also observe aggregate purchase data for each time period. Second, we assume that the panel accurately represents $q\%$ of the population. For example, if the projected market size were 100,000 and the panel size were 2000, the assumption that the panel represents 80% of the population implies that the demand estimates from the panel represent the quantity demanded by 80,000 of the 100,000 in the population.

We incorporate cross-sectional identification to develop bounds on the population distribution of preferences using the aggregate data and panel data together in each time period. Since the panel represents the same $q\%$ of the population in each time period, once we account for the information learned from the panel in the aggregate data, the remaining $(100 - q)\%$ of the population is the same over each time period. Partition the overall population Ψ into two sets: Γ, the portion represented by the panel, and Υ, the portion not represented by the panel ($\Gamma \cup \Upsilon \equiv \Psi$). Our identification proceeds in two steps. First, we use one of the four models developed in Section 2 to partially identify the distribution of preferences in the panel sample. Next, we represent the purchases made by Γ with the data from the panel, proportionally scaled up, and construct an observation each period for aggregate purchases made by the part of the population not represented by the panel. If $Q_t(\Psi)$ is the vector of purchases in period t for the entire population and $Q_t(\Gamma)$ is the vector of purchases for the population represented by the panel, the aggregate purchase observation for Υ in each t is:

$$Q_t(\Upsilon) = Q_t(\Psi) - Q_t(\Gamma). \tag{19}$$

To place bounds on demand parameters for the entire population, we combine the bounds on preferences derived from the panel data with a bound on the aggregate preferences for the remaining population derived from observing $Q_t(\Upsilon)$ over all time periods. Once we construct the residual observation $Q_t(\Upsilon)$ from aggregate and panel data, identification of the preferences of consumers in Υ is independent of the panel preference identification.

It is important to note that the assumption that the panel represents $q\%$ of the population is testable within our framework. If the bounds on the preference parameters of Υ are the empty set, then as long as we accept the assumptions on inter-temporal variations in preferences from the panel model in Section 2 that we are using, then $q\%$ is assumed to be too large. For example, this would be the case if $Q_t(\Upsilon)$ ever has any negative entries. This is a one-sided test since the data will never reveal that q is too low.

To partially identify the distribution of preferences in Υ we construct tightest bounds from the series of observations $(Q_t(\Upsilon), p_t)$ where p_t is the price vector for each t. We use two theoretical restrictions that must be satisfied by the population Υ. The first concerns dominated price movements. Here x denotes a specific product and x^- denotes all other products.

Condition 1 (*Purchase Consistency I*).

$$\forall t \neq t', \forall x, \quad Q_{xt} \geq Q_{xt'} \quad \text{if } p_{xt} \leq p_{xt'} \text{ and } p_{x^-t} \geq p_{x^-t'}. \tag{20}$$

This condition says that if the price of one product goes down and the prices of all other products go up, then we must see a higher aggregate purchase level for x in Υ.

The second condition concerns purchase behavior of the outside option x_0 relative to certain types of price changes.

Condition 2 (*Purchase Consistency II*).

$$\forall t \neq t' \quad Q_{x_0 t} \leq Q_{x_0 t'} \quad \text{if } p_t \leq p_{t'}. \tag{21}$$

This condition states that if the prices of all goods go weakly up or down, then the amount of individuals not purchasing also moves weakly up or down.

We further define the following two objects:

$$\Phi_x(p_c) \equiv p : p_x \geq p_{xc}, \qquad p_{x^-} \leq p_{x^-c}$$
$$\Delta_x(p_c) \equiv p : p_x \leq p_{xc}, \qquad p_{x^-} \geq p_{x^-c}.$$

For any counterfactual price vector p_c, $\Phi_x(p_c)$ is the set of feasible price vectors such that the price of product x is weakly greater than p_{xc} (the price of product x in the vector p_c) and the price of every other product is weakly smaller than its price in p_c. $\Delta_x(p_c)$ is the converse, where p_x is weakly lower than p_{xc} and all other products are weakly more expensive.

We can now define, given purchase consistency conditions I and II, the bounds on demand for product $x \in D$ under counterfactual price vector p_c for Υ:

$$H[P(y(D) = x)|\Upsilon] = [\max_{p_1, \ldots, p_T \in \Phi_x(p_c)} Q_{xt},$$
$$1 - \max_{p_1, \ldots, p_T \in \Delta_x(p_c)} Q_{x^-t}]. \tag{22}$$

The bounds on preferences and counterfactual demand for the residual population Υ can then be combined with those from population represented by the panel, Γ, to find the bounds for preferences and counterfactual demand for the entire population Ψ:

$$H[P(y(D) = x)|p_c] = qH[P(y(D) = x)|\Gamma, p_c]$$
$$+ (1 - q)H[P(y(D) = x)|\Upsilon, p_c]. \tag{23}$$

The simulation in the next section reveals that both the panel and cross-sectional components of this model add significant predictive power by tightening the bounds on preferences and counterfactual demand.

Simulation with cross-sectional and panel data

To simulate the scenario where we have cross-sectional and panel data, we simulate two sets of individuals. The first set represents the panelists and second set represents the consumers that are not represented in the panel. For the analysis we will observe (a) all decisions made in every time period by the panelists and (b) aggregate cross-sectional decisions (across both groups) for every time period.

In this setting we consider the preferences of the panelists to represent 80% of the entire population. The utility formulation for the panelists is exactly as in Section 2.3. For the consumers not represented by the panel, we draw the α_{i2} parameter from a uniform distribution on $[0.0, 2.0]$ and β_i independently from a uniform distribution on the range $[-2.5, -1.0]$. Therefore, these consumers are on average less price sensitive and have more varied tastes for product 2 than the panelists. We simulate 200 time periods of data and simulate each panelist's individual decisions and the aggregate decision across all consumers.

We estimate the demand curves in four steps. First, we estimate the identification region for each panelist based on their purchase data. Second, we bound the counterfactual demand for each panelist. Third, we estimate the counterfactual bounds for the aggregate consumers not represented in the panel. Fourth, we estimate the population's demand curve bounds by adding the panel and the aggregate estimates.
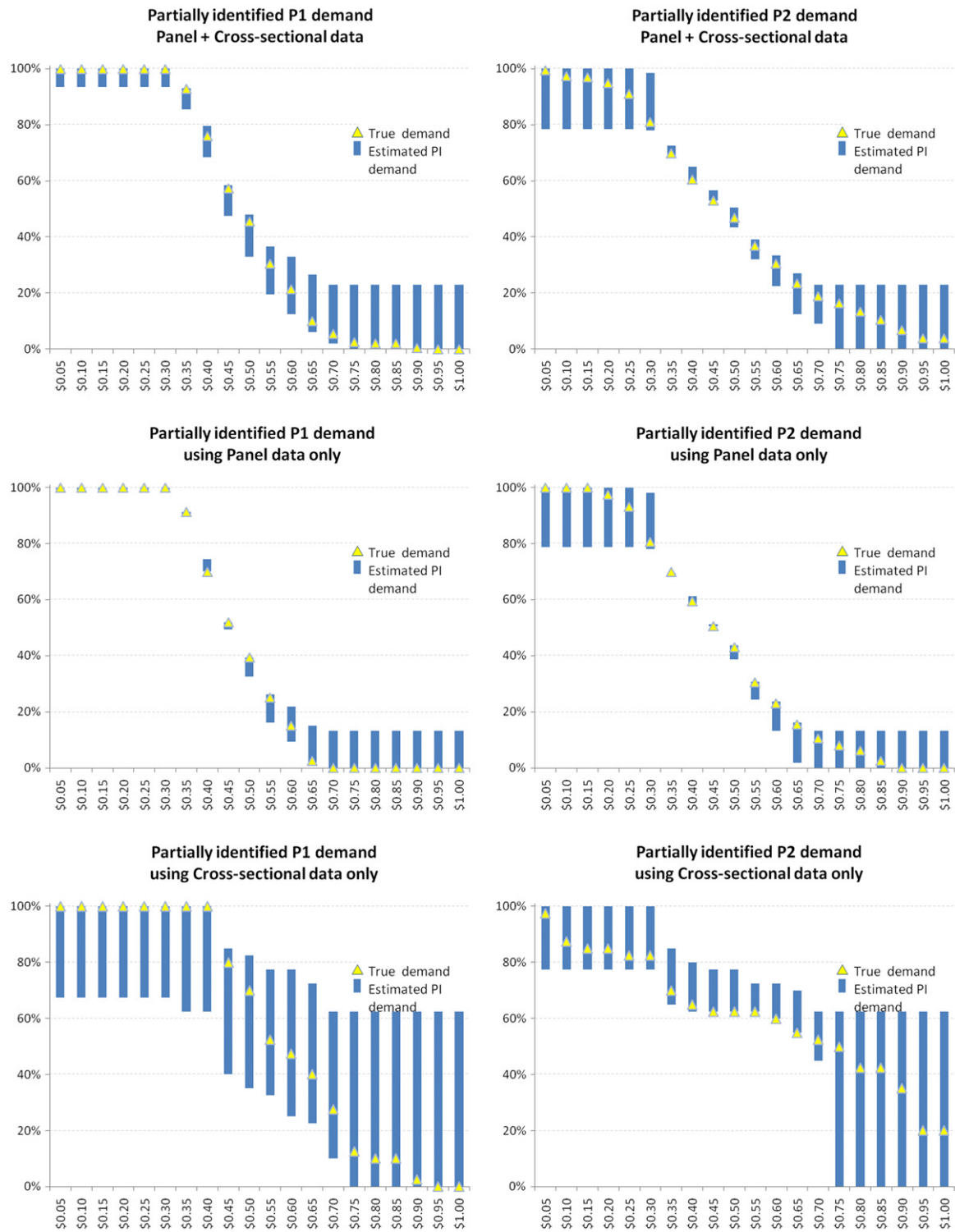
Fig. 7. Partially identified demand curves if products were sold in isolation. The rows represent the different levels of data and the columns represent the demand curves for product 1 and product 2, respectively.

Fig. 7 below displays three sets of bounds for each product's demand curve. The first row represents the demand curves using both the panel and the cross-sectional data. The second row represents the demand curve using just the panel data and the third row represents the demand curve using the cross-sectional data for the consumers not represented by the panel.

Observe that the estimated bounds with only the cross-sectional data are quite wide. This suggests that while these data do provide some information, we cannot tightly bound the counterfactual demand. On the other hand, as we have seen above, we can tightly bound the counterfactual demand with panel data. The difference between these two charts shows the additional benefit of panel data in estimating tight counterfactual bounds. We can now combine both pieces of information to create a bound for the entire population (top row of Fig. 7). In these charts observe that we get quite tight bounds in the middle of the demand curve
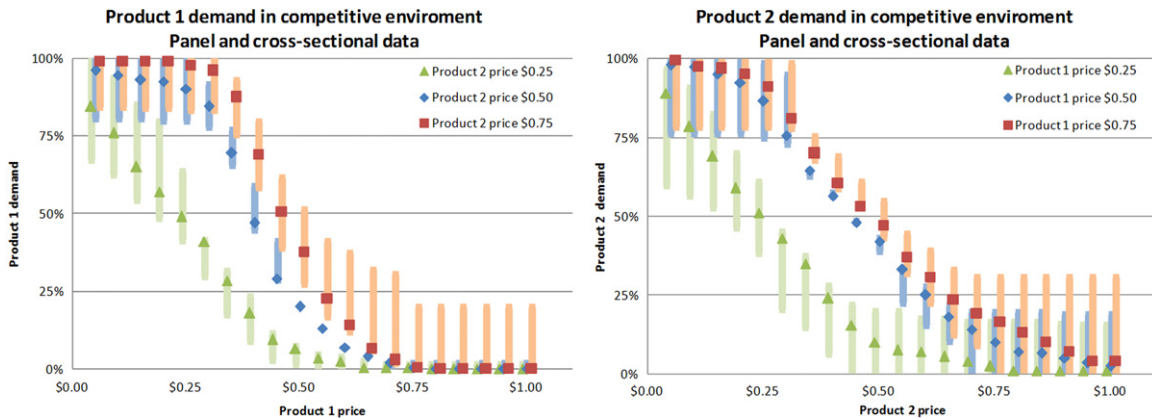
Fig. 8. Partially identified demand curves if both products were sold. The panels represent the demand curves for product 1 and product 2, respectively.
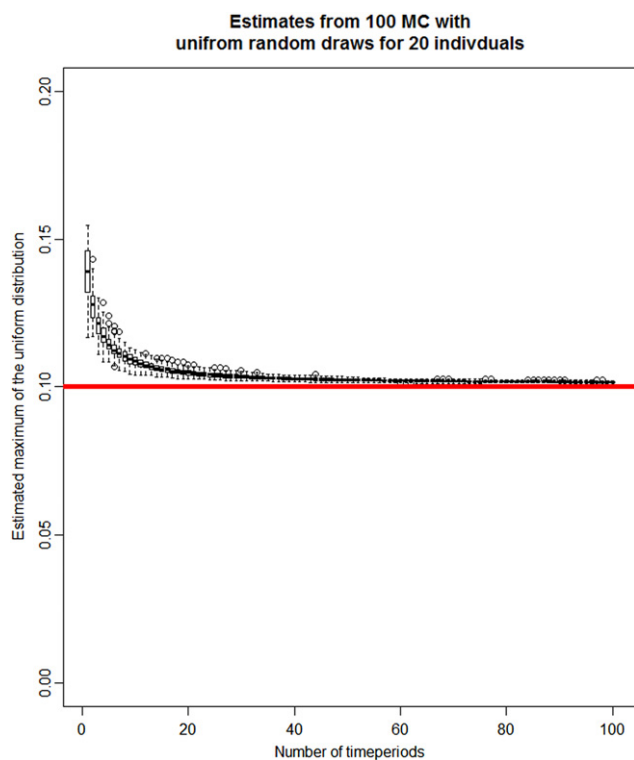

Fig. 9. Box plot of the estimated maximum of a distribution based on a small sample. Here we consider 100 experiments with 20 individuals. The experiments differ in the number of observations per consumer. The dark black lines represent the mean. The box represents the inner quartile range. The whiskers extend to the most extreme data point which is no more than 1.5 times the length of the box away from the box. The dots represent outliers that lie outside the whiskers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Table represents the percentage of estimated δ that are below the true value 0.10 by changing the number of consumers and the number of time periods. See text for more details on simulation design.

Number of consumers	Number of time periods			
	50 (%)	100 (%)	500 (%)	1000 (%)
10	77.5	32.5	15.0	2.5
20	32.5	10.0	5.0	0.0
50	25.0	0.0	0.0	0.0

two Monte Carlo studies. In the first study, we abstract from any choice model and study our estimator of the upper bound of the support from which a random variable is drawn. The study is designed as follows. Fix some number of time periods T and set the number of 'consumers' to be 20. For each of these 20 consumers, draw T realizations of δ uniformly from [0, 0.10] and compute the maximum of the T draws for each of the 20 consumers, yielding 20 maximum values of δ. Then use these 20 maximum values of δ to estimate the upper bound of the interval from which δ was drawn (the true value is 0.10) as described in Section 3.2. We then repeat this estimation procedure 100 times for each integer value of T between 1 and 100. The results are reported in Fig. 9. These suggest that we have a conservative estimate of the true maximum and as the number of realizations increase, we asymptote to the correct value.

The second Monte Carlo study we perform involves an actual choice model. Each study involves N consumers and T time periods over which choices are made. Within a consumer, a time period is distinguished by the value of δ_{nt} for that consumer n during a time period t, which is drawn uniformly from [0, 0.10]. We use the same simulation as in Section 2.1. Each time period, a consumer makes a choice according to buy one of the two goods in the market or the outside option. Then we choose the lowest value of δ for each consumer that can rationalize the choices he made during the T time periods. Then we collect these N-lowest-values-of-δ and estimate δ according to the method described in Section 3.2. We search over a grid of possible δs with 0.025 spacing. Finally, we repeat 40 times for each combination of N and T. The important statistic here is to understand how often we estimate a value of δ that is less than 0.10 as this can lead to a bias in our discrete choice estimates. The results are reported in the Table 3, which shows that for reasonable numbers of consumers or time periods our method performs well. In Fig. 10 we report the box plot for the estimates of δ from observing 50 consumers. Once again these do suggest that we have a conservative estimate of the true maximum and as the number of time periods increase we observe less variance across simulations.
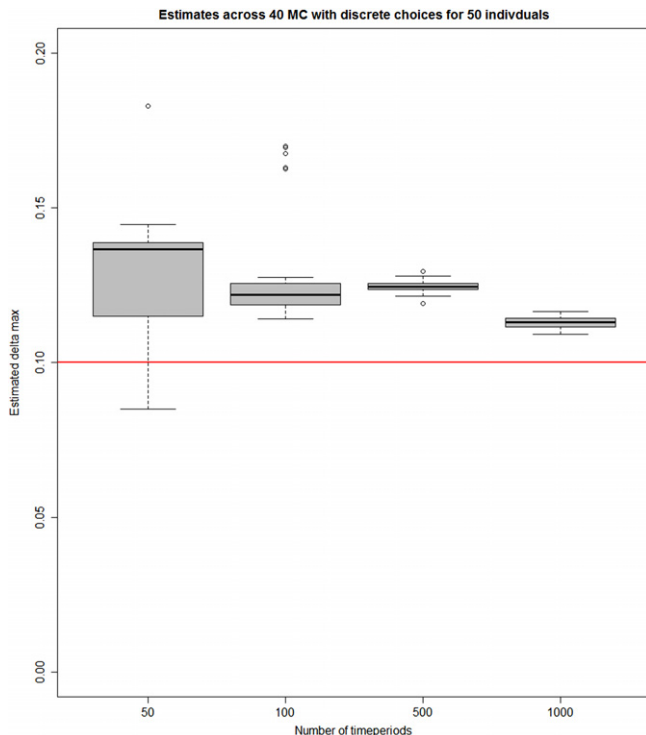
for prices between 0.35 and 0.65. However the bounds are quite wide for higher price points. This is mainly driven by the fact that we estimate wide bounds for the cross-sectional data at high price points. Overall this demand curve can be informative and can be used for firm decision-making. In Fig. 8 we display the joint demand curves when both products are sold.

Appendix B. Monte Carlo studies of the procedure to estimate δ: for online publication

In this section we study the performance of our estimator of δ, the bound on consumers' utility shocks. We do this through

Fig. 10. Box plot of the estimated values of δ after bias correction in a choice model setting. The box plot represents the estimates of δ across 40 Monte Carlo experiments with 40 observations. The experiments vary by the number of data points observed for each consumer (either 50, 100, 500 or 1000). The dark black lines represent the mean. The box represents the inner quartile range. The whiskers extend to the most extreme data point which is no more than 1.5 times the length of the box away from the box. The dots represent outliers that lie outside the whiskers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Abe, M., 1995. Nonparametric density estimation method for brand choice using scanner data. Marketing Science 14 (3), 300–325.

Bajari, P., Benkard, L., 2003. Discrete Choice Models as Structural Models of Demand: Economic Implications of Common Approaches, Working Paper.

Ben-Avika, M., McFadden, D., Abe, M., Bockenholt, U., Bolduc, D., Gopinath, D., Morikawa, T., Ramaswamy, V., Rao, V., Revelt, D., Steinberg, D., 1997. Modeling methods for discrete choice analysis. Marketing Letters 8 (3), 273–286.

Bergemann, D., Morris, S., 2005. Robust mechanism design. Econometrica 73 (6), 1771–1813.

Bergemann, D., Schlag, K., 2007. Robust Monopoly Pricing, Cowles Foundation Discussion Papers 1527.

Bergemann, D., Schlag, K., 2008. Pricing without priors. Journal of the European Economic Association 6, 560–569.

Berry, S., Haile, P., 2009. Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers, Working Paper.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.

Bloch, D., Gastwirth, J., 2002. On a simple estimate of the reciprocal of the density function. Annals of Mathematical Statistics 39, 1083–1085.

Blundell, R., Browning, M., Crawford, I., 2003. Nonparametric engel curves and revealed preferences. Econometrica 71 (1), 205–240.

Blundell, R., Browning, M., Crawford, I., 2008. Best nonparametric bounds on demand responses. Econometrica 76 (6), 1227–1262.

Bronnenberg, B., Kruger, M., Mela, C., 2008. The IRI marketing data set. Marketing Science 27 (4), 745–748.

Dreze, X., Hoch, S., Purk, M., 1994. Shelf management and space elasticity. Journal of Retailing 70 (4), 301–326.

Einav, L., Leibtag, E., Nevo, A., 2010. Recording discrepancies in nielsen homescan data: are they present and do they matter? Quantitative Marketing and Economics 8 (2), 207–239.

Erdem, T., Imai, S., Keane, M., 2003. Brand and quantity choice dynamics under price uncertainty. Quantitative Marketing and Economics 1 (1), 5–64.

Erdem, T., Keane, M., 1996. Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. Quantitative Marketing and Economics 15 (1), 1–20.

Erdem, T., Keane, M., Sun, B., 1999. Missing price and coupon availability data in scanner panels: correcting for the self-selection bias in choice model parameters. Journal of Econometrics 89, 177–196.

Fiebig, D., Keane, M., Louviere, J., Wasi, N., 2010. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. Marketing Science 29 (3), 393–421.

Fong, N., Simester, D., Anderson, E., 2011. Private Label vs National Brand Price Sensitivity: Evaluating Non-experimental Identification Strategies, Working Paper.

Geweke, J., 2012. Nonparametric Bayesian modeling of monotone preferences for discrete choice experiments. Journal of Econometrics 171 (2), 185–204.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18 (1), 141–153.

Haistie, T., Tibshirani, R., 1990. Generalized Additive Models. CRC Press.

Hall, P., Park, B.U., 2002. New methods for Bias correction at endpoints and boundaries. The Annals of Statistics 30 (5), 1460–1479.

Hendel, I., Nevo, A., 2006. Measuring the implications of sales and consumer inventory behavior. Econometrica 74 (6), 1637–1673.

Horowitz, J., Manski, C., 1995. Identification and robustness with contaminated and corrupted data. Econometrica 63 (2), 281–302.

Karunamuni, R., Alberts, T., 2005. On boundary correction in Kernel density estimation. Statistical Methodology 2 (3), 191–212.

Keane, M., 1997. Modeling heterogeneity and state dependence in consumer choice behavior. Journal of Business and Economic Statistics 15 (3), 310–327.

Keane, M., Sauer, R., 2009. Classification error in dynamic discrete choice models: implications for female labor supply behavior. Econometrica 77 (3), 975–991.

Keane, M., Sauer, R., 2010. A computationally practical simulation estimation algorithm for dynamic panel data models with unobserved endogenous state variables. International Economic Review 51 (4), 925–958.

Khan, R., Misra, K., Singh, V., 2012. Will a Fat Tax Work? Working Paper, London Business School.

Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157.

Manski, C., 1975. Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3 (3), 205–228.

Manski, C., 2005. Social Choice with Partial Knowledge of Treatment Response. Princeton University Press, Princeton, NJ.

Manski, C., 2007. Partial identification of counterfactual choice probabilities. International Economic Review 48 (4), 1393–1410.

Matzkin, R., 1993. Nonparametric identification and estimation of polychotomous choice models. Journal of Econometrics 58 (1–2), 137–169.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior Frontiers in Econometrics. Academic Press, New York, NY.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15, 447–470.

Nielsen, 2009. The Nielsen Company and Information Resources, Inc. Form Production Joint Venture To Recruit and Maintain Panelists for Consumer Shopping Panels. http://www.nielsen.com/us/en/insights/press-room/2009/nielsen-iri-joint.html.

Renou, L., Schlag, K., 2010. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. Journal of Economic Theory 145, 264–286.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. The Review of Economics and Statistics 80 (4), 647–657.

Samuelson, P., 1938. The numerical representation of ordered classifications and the concept of utility. Review of Economic Studies 6 (1), 65–70.

Samuelson, P., 1948. Consumption theory in terms of revealed preference. Economica 15 (60), 243–253.

Savage, L., 1951. The theory of statistical decisionbel. Journal of the American Statistical Association 46, 55–67.

Siddiqui, M., 1960. Distribution of quantiles in samples from a bivariate population. Journal of Research of the National Bureau of Standards 64B, 145–150.

SymphonyIRI,, 2012. Consumer Network: Providing Sales and Marketing Professionals a Clear Picture of Consumer Behavior, Data Sheet.

Varian, H., 1982. The nonparametric approach to demand analysis. Econometrica 50 (4), 945–974.

Varian, H., 1983. Non-parametric tests of consumer behaviour. Review of Economic Studies 50 (1), 99–110.