Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Market Discovery Algorithm to Estimate a General Class of Nonparametric Choice Models

Garrett van Ryzin, Gustavo Vulcano

MANAGEMENT SCIENCE
Vol. 61, No. 2, February 2015, pp. 281–300
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2040

A Market Discovery Algorithm to Estimate a General Class of Nonparametric Choice Models

Garrett van Ryzin
Graduate School of Business, Columbia University, New York, New York 10027, gjv1@columbia.edu

Gustavo Vulcano
Leonard N. Stern School of Business, New York University, New York, New York 10012, gvulcano@stern.nyu.edu

We propose an approach for estimating customer preferences for a set of substitutable products using only sales transactions and product availability data. The underlying demand framework combines a general, nonparametric discrete choice model with a Bernoulli process of arrivals over time. The choice model is defined by a discrete probability mass function (pmf) on a set of possible preference rankings of alternatives, and it is compatible with any random utility model. An arriving customer is assumed to purchase the available option that ranks highest in her preference list. The problem we address is how to jointly estimate the arrival rate and the pmf of the rank-based choice model under a maximum likelihood criterion. Since the potential number of customer types is factorial, we propose a *market discovery* algorithm that starts with a parsimonious set of types and enlarge it by automatically generating new types that increase the likelihood value. Numerical experiments confirm the potential of our proposal. For a realistic data set in the hospitality industry, our approach improves the root mean square errors between predicted and observed purchases computed under independent demand model estimates by 67%–93%.

Keywords: demand estimation; random utility models; choice behavior; demand untruncation; column generation
History: Received January 4, 2012; accepted June 21, 2014, by Serguei Netessine, operations management. Published online in *Articles in Advance* October 29, 2014.

1. Introduction

Demand estimation is a central task in retail operations and revenue management (RM). For simplicity, many demand models in practice assume independent demand for each product. However, if products are substitutes, then the demand for a given product will be a function of the set of alternatives available to consumers when they make their purchase decisions (the *offer set*). Such choice behavior is a topic of great interest among revenue management researchers and practitioners alike because it has obvious and significant revenue consequences.

Estimating demand is particularly challenging when product availability varies over time as a result of either inventory stockouts or deliberate availability controls applied by the seller. Two consequences of changing availability are the occurrence of *spilled* and *recaptured* demand. The former refers to the demand lost when a consumer's first choice is unavailable. Recapture refers to demand for a product that results when customers substitute because their preferred product is unavailable. Both behaviors can complicate demand estimation.

Spilled demand could potentially be observed if a seller were to record initial customer requests; however, this is not common. In many physical retail stores, customer shopping is self-service (buying off a shelf), and there is no realistic opportunity for a seller to elicit customer preferences. Even when customers interact with a sales agent, other issues make eliciting preferences difficult. For instance, in the case of hotel operations, Queenan et al. (2007) mention multiple availability inquiries from the same customer, incorrect categorization of turndowns by reservations agents, and customer requests that arrive through a channel not controlled by the firm. Clearly, if historical sales data were left uncorrected for spill, the true underlying demand would be underestimated. This bias contributes to the well-known *spiral-down* phenomenon, in which a firm's expected revenue decreases over time (see Cooper et al. 2006) because of a cycle of lower estimates of demand leading to progressively lower availability. Correcting this bias is important for preventing such spiral-down problems.

Recaptured demand due to substitution, in contrast, produces an increase in observed sales of alternative available products. Ignoring this effect leads to an overestimation bias among the products that are available. For example, in retailing, customers may buy an alternative color, size, or flavor if their first preference is out of stock. A similar phenomenon occurs among airline passengers who may buy a ticket on an alternate flight with a different departure time or routing. Empirical studies from different industries confirm that such substitution is significant. Gruen et al. (2002) report recapture rates of 45% across eight categories at retailers worldwide, and for airline passengers recapture rates are acknowledged to be in the range of 15%–55% (Ja et al. 2001).

Estimating customer choice behavior first requires specifying a choice model, either parametric or nonparametric. Parametric models assume a specific functional form relating choice alternatives and their attributes to purchase probabilities. Given such a form, one then estimates its parameters from data. One widely used example of a parametric model is the multinomial logit model (MNL) (e.g., see Ben-Akiva and Lerman 1994, Train 2003). A convenient aspect of the MNL model is that the likelihood of purchase can be readily recalculated if the mix of available related products changes (e.g., because of items being sold out or restocked). Yet a major disadvantage is that the MNL exhibits the independence from irrelevant alternatives (IIA) property, which can lead to unrealistic substitution patterns in cases where alternatives have shared attributes, despite promising results reported in the literature (e.g., see Ratliff et al. 2008). Other parametric models such as the nested logit or mixed logit help overcome this IIA property (again, see Ben-Akiva and Lerman 1994, Train 2003). The greatest advantage of parametric models is the ability to include covariates, such as product features and price, that can help explain consumer preferences for alternatives. This also enables parametric models to extrapolate choice predictions to new alternatives that have not been observed in historical data and to predict how changes in product attributes such as price affect choice outcomes. Still, the drawback of any parametric model is that one must make assumptions about the structure of preferences and the relevant covariates that influence it, which requires expert judgement and trial-and-error testing to determine an appropriate specification.

There is also a balance between specification and estimation error when selecting a parametric model. A complex model with many attributes, nests, and latent classes may be able to better approximate a wide range of choice behavior and reduce the specification error. However, estimating the parameters of such a model from a finite set of data may lead to significant estimation error. A more parsimonious model, although potentially less faithful in terms of representing underlying choice behavior, is less prone to estimation error. The overall accuracy of a model in predicting demand requires the modeler to balance specification and estimation errors.

In this paper, we take a nonparametric approach to this problem. We use a simple, though quite general, nonparametric choice model in which customer types are defined by their rank ordering of all alternatives (along with the no-purchase alternative). When faced with a choice from an offer set, a customer is assumed to purchase the available product that ranks highest in her preference list—or she does not purchase at all if the no-purchase alternative ranks higher than any available product. With a finite number of alternatives, there is a finite number of rankings and hence a finite number of customer types. Demand is then described by a discrete probability mass function (pmf) on the set of customer types. This type of rank-based choice model has previously been applied in economics and psychology (see Block and Marschak 1960, Manksi 1977, Falmagne 1978, Barberá and Pattanaik 1986). One of the first applications in operational settings was the retail assortment problem studied by (Mahajan and van Ryzin 2001). As they point out, several common demand processes studied in the literature can be modeled as special cases of a rank-based choice model (e.g., MNL, Markovian second choice, universal backup, Lancaster demand, the independent demand model).

Of course, the potential number of preference lists (customer types) in this rank-based choice model is factorial in the number of alternatives. As a result, it is well understood in the literature that this model suffers from nonidentifiability (i.e., two different pmfs may explain the same observed data; see Sher et al. 2011, Theorem 5). Indeed, even if we had observation of transactions for all possible offer sets, it is not possible to identify the customer type distribution when there are four or more alternative products in the rankings since the number of rankings exceeds the number of offer sets. To circumvent this problem, *partial identification* approaches have been proposed in the literature: (Manksi 2007) explores bounds on choice probabilities under counterfactual choice situations, whereas (Sher et al. 2011) focus on identifying the pmf on the preference lists, finding maximum and minimum probabilities of any ranking consistent with observed data. Our approach to selecting rankings is to start with a simple, parsimonious set of rankings and progressively augment this initial set with new types that increase the likelihood.

Specifying the subset of relevant types to use in a rank-based choice model is analogous to the problem of specifying the form of model used in the parametric case and involves similar trade-offs; using the complete set of possible customer types provides maximum flexibility in modeling the choice behavior of a given population of customers, but the large number of parameters involved can lead to significant estimation error (overfitting). In contrast, a more parsimonious set of customer types, although more restrictive in terms of representing choice behavior, will have fewer parameters to estimate and hence is less prone to estimation error.

In this paper, we focus on two problem specifications. In the first, we assume that the modeler can distinguish a period with no arrival from a period where an arrival ended in a no purchase. This specification is applicable to online retailers that keep track of customer visits, purchases, and no purchases. The unobservable data are the type of the customer who arrives in a given period. The problem we address is how to estimate the pmf of a rank-based choice model that best explains a set of observed transactions. The only required inputs are observed historical sales, product availability data, and an initial set of customer types. We then develop a market discovery algorithm for automatically augmenting this initial set of types. Every iteration of this constructive procedure consists of two stages: (1) given the set of types under consideration, we estimate the pmf to maximize the likelihood function, and (2) we add a new type to the current set in order to improve the log-likelihood value. To this end, we use dual information from the maximum likelihood estimation problem to solve a large-scale concave program via column generation. We use duality theory to identify conditions for the entering column (i.e., new customer type) that ensure it is a direction of improvement for the likelihood function, and we control for overfitting at the moment of incrementing the customer type set by checking the improvement in the log-likelihood value.

The second specification is applicable to brick-and-mortar retailers, where not all customer visits are tracked and therefore it is not possible to distinguish between a period with no arrival and a period with an arriving customer who decides not to purchase. The same problem arises in online retailers that record transactions rather than visits. Our procedure extends the arguments described for the previous variation and corrects for both sources of incompleteness in the data: (a) the arrival (or not) of a customer and (b) the type of arriving customer.

Our numerical experiments for both variants of the problem show that after adding only a handful of new types through our column generation approach, the likelihood function increases rapidly and exhibits only marginal improvements thereafter, confirming the practical potential of our procedure to reduce the inherent complexity of the rank-based choice model estimation.

The remainder of this paper is organized as follows. In §2, we review the related literature. In §3 we introduce the demand model. Our estimation procedure for the uncensored demand case is described in §4, followed in §5 by numerical experiments based on both synthetic and real-world data. In §6 we explain how to extend the ideas to the censored demand case. Finally, we present our conclusions in §7. All proofs are included in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mnsc.2014.2040).

2. Literature Review

Since the late 1990s, there has been growing interest in the retail operations and marketing science literature on estimating choice behavior and lost sales. (Anupindi et al. 1998) present a method for estimating consumer demand when the first-choice variant is not available. They assume a continuous-time model of demand and develop an expectation–maximization (EM) method to uncensor times of stockouts for a periodic review policy, with the constraint that at most two products stock out in order to handle a manageable number of variables. The authors find maximum likelihood estimates of arrival rates and substitution probabilities. Other papers in this area are Swait and Erdem (2002), Borle et al. (2005), Chintagunta and Dubé (2005), Kalyanam et al. (2007), Kök and Fisher (2007), Bruno and Vilcassim (2008), and more recently, Musalem et al. (2010).

Interest in estimation and forecasting techniques has also been growing in the RM research literature, which traditionally focused more on optimization of availability controls. Queenan et al. (2007) review unconstraining methods for univariate demand models that are used commonly in RM practice. Talluri and van Ryzin (2004, §5) develop an EM method to jointly estimate arrival rates and parameters of a MNL choice model based on consumer-level panel data under unobservable no purchases. More recently, Newman et al. (2014) present a heuristic that accelerates the computational times of Talluri and van Ryzin (2004) but leads to estimates that are not based on maximum likelihood criteria.

Ratliff et al. (2008) provide a comprehensive review of the demand untruncation and choice behavior estimation literature in RM. They also propose a heuristic to jointly estimate spill and recapture across numerous flight classes, by using balance equations that generalize the proposal of Andersson (1998). A similar approach was presented before by Ja et al. (2001). Talluri (2009) proposes an estimation method that relies on a finite-population model and a heuristic that exploits the functional form of the MNL model, the variety of offer sets in a typical RM setting, and qualitative knowledge of arrival rates. He also provides experiments on synthetic data that show some promising features of the resulting estimates. In Vulcano et al. (2012), we analyze a model of demand that combines a MNL choice model with nonhomogeneous Bernoulli arrivals over multiple periods. The problem addressed was how to jointly estimate the preference weights of the products and the arrival rates of customers when the availability of products changes over time. The key idea was to view the problem in terms of primary (or first-choice) demand and to treat the observed sales as incomplete observations of such primary demand. By contrast, in our new proposal, we work on individual purchase records rather than on demand data aggregated over long periods (e.g., a day or a week). But the main difference in our work here is the generality of the choice model; indeed, the MNL model is just a special case of our nonparametric choice model.

Haensel and Koole (2011) present an EM-based procedure to estimate a nonparametric choice model that is quite similar to ours, consisting also of a set of customer types that are defined by preference orderings. However, they do not explicitly take into account the incompleteness of the data with respect to arrival and no-arrival periods and instead use extrapolation heuristics to fill in for the lack of observations in a given time interval. Their work does not address the important issue of how to efficiently construct the set of customer types used.

From a choice modeling point of view, our paper is closest to Farias et al. (2013), who also propose procedures for estimating a nonparametric choice model such as ours. However, the approach to estimation is different. Whereas we find maximum likelihood estimates of the choice demand model, Farias et al. (2013) use a robust approach, finding the distribution over customer types that produces the worst-case revenue compatible with the observed data for a given fixed assortment. These robust revenues are computed approximately using constraint sampling. They then show that the demand model constructed by the robust approach is approximately the *sparsest-choice model*, where sparsity is measured by the number of customer types that occur with positive probability in the population, and that the sparsity grows with the amount of observed data. Farias et al. address the demand untruncation in their real-data-based numerical experiments in an ad hoc way, using a simple proportionality method discussed in Queenan et al. (2007). Another related model in the context of assortment pricing is the rank pricing model of Rusmevichientong et al. (2006), but their approach requires access to samples of entire customer preference lists that are unlikely to be available in real-world applications.

One of the main goals of choice modeling in general is to use the inferred estimates as inputs for revenue optimization routines. In the context of rank-based choice models, such optimization routines were explored by Mahajan and van Ryzin (2001), Honhon et al. (2012), and Farias et al. (2012) for retail assortment, and by Zhang and Cooper (2006), van Ryzin and Vulcano (2008), Chen and Homem-de-Mello (2010), Chaneton and Vulcano (2011), and Kunnumkal (2014) for airline RM.

3. Problem Description

3.1. Model Basics

A set of n substitutable products is sold over T periods indexed by $t = 1, 2, \ldots, T$. The full set of products is denoted $\mathcal{N} = \{0, 1, \ldots, n\}$, $n \geq 1$, where 0 stands for the outside or no-purchase option. Customers arrive according to a discrete-time, homogeneous Bernoulli arrival process, where in each time period t an arrival occurs with probability $0 < \lambda < 1$, which we call the *arrival rate*. If $\lambda \ll 1$, this arrival process can be considered a discrete-time approximation to a Poisson arrival process. To avoid a limiting and noninteresting case, we assume there is at least one period with an observed transaction. In cases where the arrival rate may not be homogeneous throughout the selling horizon,[1] one can partition time periods into multiple segments in such a way that within each segment the arrival rate is constant. For ease of exposition, however, we consider only the case where the arrival rate is the same in all time periods.

Customers are assumed to have a rank-based preference for products. That is, each customer has a preference list (or total ranking), σ, of the products in $\mathcal{N}$. Each preference list σ defines a *customer type*. A customer of type σ prefers product h to product j if and only if $\sigma(h) < \sigma(j)$. This preference extends to the no-purchase alternative as well, since the ranking includes alternative 0. Note that the meaningful rankings are only the ones truncated at the position of product 0; that is, a preference list σ will have as many elements as $\sigma(0) - 1$, where we will assume that $\sigma(0) \geq 2$, since a customer type whose first preference is to purchase nothing is irrelevant. Considering all such meaningful rankings, the number of customer types in the market is at most $K = \sum_{i=1}^{n} \binom{n}{i} i!$; i.e., $K = O(n!)$.

Arriving customers are assumed to be of type $\sigma^{(i)}$ with probability $x_i = \mathbb{P}(\sigma^{(i)})$, $i = 1, \ldots, K$. This pmf for customer types is denoted by $\mathbf{x} = \mathbb{P}(\boldsymbol{\sigma})$ and must be estimated from the data. The random draw of customer types is assumed to be time homogeneous and independent across periods. Time homogeneity can be relaxed by specifying different ranking models for different intervals of time (e.g., early- versus late-arriving customers), but to simplify the exposition, we focus only on the time-homogeneous case.

In each period, a set of products $S_t \subset \mathcal{N}$ is offered, with $0 \in S_t$; i.e., the no-purchase alternative is always

[1] For some empirical evidence in the airline industry, we refer the reader to Haensel and Koole (2011).

available. We assume that $|S_t| \geq 2$, so that at least one product other than the no-purchase option is available in each period; otherwise, if $S_t = \{0\}$, then the period is not informative and can be dropped from the data set. An arriving customer chooses her most preferred product among those available in S_t. In symbols, if there is an arrival of type σ in period t, that customer chooses $j_t = \arg\min_{j \in S_t} \sigma(j)$.

We further assume that the arrivals and no arrivals are observed, and that the no-purchase outcome of an arrival is also observed. This corresponds, for example, to the case of online retailers who record shopping data (i.e., customer visits to a website that may end up in either a transaction or a no purchase). This variant allows us to more easily explain our method for estimating which customer types are significant and in which proportion they appear in the market. In §6, we consider the effect of censoring when no-purchase outcomes are unobservable. The extra difficulty there is that one cannot distinguish a period with no arrival from a period with an arrival that did not purchase because of a limited availability of alternatives.

A key observation is that, in general, it is not possible to precisely identify customer types from transactions data, since for each period with an observed purchase the most we can say is that the observed choice is *compatible* with some subset of the possible types (i.e., those types who rank the purchased product higher than all other available products). For each period with an arrival and observed no purchase, we can only say that the customer who arrived preferred the no-purchase alternative to any of the products offered. Hence, sales data provide only incomplete observations of the choice model we wish to estimate. More formally, for the full set of types $\{\sigma^{(1)}, \ldots, \sigma^{(K)}\}$ and $j_t \in S_t$, define the set

$$\mathcal{M}_t(j_t, S_t) = \{i\colon \sigma^{(i)}(j_t) < \sigma^{(i)}(k) \quad \forall\, k \in S_t, k \neq j_t\}.$$

In other words, $\mathcal{M}_t(j_t, S_t)$ is the set of customer types for which product j_t (i.e., the product in S_t chosen in period t) is ranked highest among the available products in S_t. The definition also covers the case $j_t = 0$ and the corresponding set $\mathcal{M}_t(0, S_t)$. Hence, $\mathcal{M}_t(j_t, S_t)$ is the set of customer types that is compatible with the observed transaction in period t.

Given a probability distribution $\mathbf{x}$ over the distribution of types, the probability that a random arrival chooses product j_t when set S_t is offered is given by

$$\mathbb{P}(j_t \mid S_t, \mathbf{x}) = \sum_{i \in \mathcal{M}_t(j_t, S_t)} x_i \quad \text{if } j_t \in S_t, \tag{1}$$

and $\mathbb{P}(j_t \mid S_t, \mathbf{x}) = 0$ if $j_t \notin S_t$. When the dependence over the distribution of types is clear, we will just write $\mathbb{P}(j_t \mid S_t)$. We will assume, without loss of generality, that $\mathbb{P}(j_t \mid S_t) > 0$ for all $j_t \in S_t$ (otherwise, product j_t could be dropped from S_t).

The following running example illustrates the model and related definitions.

Illustrative Example. The selling horizon has $T = 10$ periods, and there are $n = 5$ products (plus the always-available no-purchase alternative 0). For simplicity, assume for now that the market consists of only two relevant customer types with rank preferences $\sigma^{(1)} = (1, 2)$ and $\sigma^{(2)} = (1, 2, 3, 4, 5)$, respectively.[2] For example, if product indices follow an increasing price order, type 1 customers (price-sensitive customers) prefer only the cheapest products 1 and 2, whereas type 2 customers (price-insensitive customers) are willing to buy up to the most expensive product 5.

Table 1 describes the parameters of our model. A label "Yes" ("No") in position (j, t) means that product j is available (unavailable) in period t. The table also reports the observed sales (a zero stands for no purchase). Following this information, the sets $\mathcal{M}_t(j_t, S_t)$ of compatible customer types for each period that could be inferred from the previous data are given. Finally, the table lists the unobservable data: the specific type of the customer that arrived in each period (or a no-arrival event).

For instance, in period 1, product 1 was purchased. The offer set was $S_1 = \{0, 1, 2, 3, 4\}$. Since product 1 is the top choice for both customer types, we cannot distinguish which type indeed arrived and purchased the product; i.e., $\mathcal{M}_1(1, \{0, 1, 2, 3, 4\}) = \{1, 2\}$. It turned out that in this period the true (unobserved) arrival was of type 1.

In the second period, the full assortment was available, but no arrival occurred. This period is uninformative for the purpose of estimating preferences and therefore disregarded. In period 3, product 4 was purchased. The only compatible type with this outcome is type 2, since it is the only type who has product 4 ranked higher than not purchasing and ranked higher than product 5, i.e., with $\sigma^{(2)}(4) < \sigma^{(2)}(0)$ and $\sigma^{(2)}(4) < \sigma^{(2)}(5)$. Hence, $\mathcal{M}_3(4, \{0, 4, 5\}) = \{2\}$.

Period 8 shows yet a different situation. Product 3 is the only one available, and an arrival occurred but no transaction was observed. This is only possible as a result of a type 1 arrival. Note that we could not have had an arrival of a type 2 customer since she would have purchased the product. So $\mathcal{M}_8(0, \{0, 3\}) = \{1\}$.

3.2. Formulation of the Estimation Problem

Let $\mathcal{P}$ be the set of periods with purchases. Let $\bar{\mathcal{P}}_\lambda$ be the set of periods with arrivals that trigger no purchases, and let $\bar{\mathcal{P}}_{\bar{\lambda}}$ be the set of periods with no

[2] In fact, for $n = 5$, there are really $K = 325$ possible customer types.

Table 1 Availabilities and Purchases for the Illustrative Example

Observable data: Availabilities and transactions										
	Period									
Product	1	2	3	4	5	6	7	8	9	10
1	Yes	Yes	No	Yes	No	No	No	No	Yes	No
2	Yes	Yes	No	No	No	Yes	No	No	Yes	Yes
3	Yes	Yes	No	No	Yes	Yes	No	Yes	No	Yes
4	Yes	Yes	Yes	No	No	No	No	No	Yes	No
5	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Transactions	1	No arrival	4	1	3	2	5	0	1	No arrival
Compatible customer types										
Sets $\mathscr{M}_t(j_t, S_t)$	{1, 2}	—	{2}	{1, 2}	{2}	{1, 2}	{2}	{1}	{1, 2}	—
Unobservable data										
Customer types	1	—	2	1	2	1	2	1	2	—

arrivals, with $T = |\mathscr{P}| + |\bar{\mathscr{P}}_\lambda| + |\bar{\mathscr{P}}_{\bar{\lambda}}|$. The incomplete data log-likelihood function is given by

$$\begin{aligned}\mathscr{L}_I(\mathbf{x}, \lambda) &= \sum_{t\in\mathscr{P}}(\log\lambda + \log\mathbb{P}(j_t \mid S_t)) \\ &\quad + \sum_{t\in\bar{\mathscr{P}}_\lambda}\big(\log\lambda + \log\mathbb{P}(0 \mid S_t)\big) + \sum_{t\in\bar{\mathscr{P}}_{\bar{\lambda}}}\log(1-\lambda) \\ &= \sum_{t\in\mathscr{P}}\bigg(\log\lambda + \log\bigg(\sum_{i\in\mathscr{M}_t(j_t, S_t)} x_i\bigg)\bigg) \\ &\quad + \sum_{t\in\bar{\mathscr{P}}_\lambda}\bigg(\log\lambda + \log\bigg(\sum_{i\in\mathscr{M}_t(0, S_t)} x_i\bigg)\bigg) \\ &\quad + \sum_{t\in\bar{\mathscr{P}}_{\bar{\lambda}}}\log(1-\lambda).\end{aligned}$$

The first term accounts for the likelihood of the observed transactions in the periods with purchases (e.g., periods 1 and 3–9 in Table 1). The second term accounts for the no-purchase periods, where an arriving customer preferred not to buy (e.g., as in period 8 in Table 1). The third term accounts for the periods with no arrivals (e.g., periods 2 and 10 in Table 1). A more compact representation of the log-likelihood function is

$$\begin{aligned}\mathscr{L}_I(\mathbf{x}, \lambda) &= \sum_{t\in\mathscr{P}}\log\bigg(\sum_{i\in\mathscr{M}_t(j_t, S_t)} x_i\bigg) + \sum_{t\in\bar{\mathscr{P}}_\lambda}\log\bigg(\sum_{i\in\mathscr{M}_t(0, S_t)} x_i\bigg) \\ &\quad + (|\mathscr{P}| + |\bar{\mathscr{P}}_\lambda|)\log\lambda + |\bar{\mathscr{P}}_{\bar{\lambda}}|\log(1-\lambda). \quad (2)\end{aligned}$$

The function is separable in $\mathbf{x}$ and λ and globally concave in $(\mathbf{x}, \lambda)$. The maximizer λ^* is unique and has a closed form given by $\lambda^* = (|\mathscr{P}| + |\bar{\mathscr{P}}_\lambda|)/T$.[3] To simplify notation, we define $\hat{T} = |\mathscr{P}| + |\bar{\mathscr{P}}_\lambda|$, the number of periods where arrivals occurred. In what follows, assume that the arrival rate λ^* has already been established.

[3] For the example in Table 1, the maximum likelihood estimate for the arrival rate would be $\lambda^* = 0.8$.

The remaining maximum likelihood estimation (MLE) problem can be formulated as follows:

$$\begin{aligned}&\max_{\mathbf{x}\geq 0}\ \mathscr{L}_I(\mathbf{x}) \\ &\text{s.t.}\ \sum_{i=1}^K x_i = 1.\end{aligned} \quad (3)$$

Our objective is to find a solution $\mathbf{x}^*$ to problem (3).

For our analysis, we introduce new variables $\mathbf{y}$ defined as $y_t = \sum_{i\in\mathscr{M}_t(j_t, S_t)} x_i$, representing the aggregate likelihood of the customer types that would pick alternative j_t in period t, i.e., $y_t = \mathbb{P}(j_t \mid S_t)$ according to (1). Then, we rewrite the incomplete data log-likelihood function in terms of $\mathbf{y}$:

$$\mathscr{L}_I(\mathbf{y}) = \sum_{t\in\mathscr{P}}\log y_t + \sum_{t\in\bar{\mathscr{P}}_\lambda}\log y_t.$$

Define the matrix $A \in \{0,1\}^{\hat{T}\times K}$ with elements $a_{ti} = 1$ if $i \in \mathscr{M}_t(j_t, S_t)$ (here, j_t could also be zero) and $a_{ti} = 0$ otherwise. The matrix A has one row per period and one column per customer type, with $a_{ti} = 1$ when customer type i is compatible with the transaction observed in period t. We can then formulate an optimization problem equivalent to (3) in terms of the variables $(\mathbf{x}, \mathbf{y})$:

$$\begin{aligned}&\max_{\mathbf{x},\mathbf{y}\geq 0}\ \mathscr{L}_I(\mathbf{y}) \\ &\text{s.t.}\ \sum_{i=1}^K x_i - 1 = 0, \\ &\qquad A\mathbf{x} - \mathbf{y} = 0.\end{aligned} \quad (4)$$

This is a concave program with linear constraints and bounded objective $\mathscr{L}_I(\mathbf{y}) \leq 0$ for all $\mathbf{x}, \mathbf{y} \geq 0$. Take a maximum $\mathbf{x}^*$ of problem (3). Note that the point $(\mathbf{x}^*, \mathbf{y}^*)$, where $y_t^* = \sum_{i\in\mathscr{M}_t(j_t, S_t)} x_i^*$, is an optimal solution to problem (4).

3.3. Theoretical Properties of the MLE

We next analyze optimality conditions for problem (4). Define the Lagrangian dual function

$$\Theta(\beta, \boldsymbol{\mu}) = \sup_{\mathbf{x},\mathbf{y}\geq 0}\bigg\{\mathscr{L}_I(\mathbf{y}) + \beta\bigg(\sum_{i=1}^K x_i - 1\bigg) + \boldsymbol{\mu}^{\mathsf{T}}(A\mathbf{x} - \mathbf{y})\bigg\},$$

and consider the Lagrangian dual problem

$$\min_{\beta,\boldsymbol{\mu}}\Theta(\beta, \boldsymbol{\mu}), \quad \text{with } \beta \in \mathscr{R},\ \boldsymbol{\mu} \in \mathscr{R}^T. \quad (5)$$

The next result follows from the strong duality theorem (e.g., see Bazaraa et al. 2006, Theorem 6.2.4) and will be helpful for our estimation approach described later in §4.

Proposition 1. *The primal estimation problem* (4) *and its dual* (5) *have the same* (*finite*) *objective value at optimality. In particular,* $(\mathbf{x}^*, \mathbf{y}^*)$ *is an optimal solution for the primal problem, and* $(\beta^*, \boldsymbol{\mu}^*)$ *is an optimal solution for the dual problem if and only if* $(\mathbf{x}^*, \mathbf{y}^*, \beta^*, \boldsymbol{\mu}^*)$ *is a saddle point of the Lagrangian function*

$$\phi(\mathbf{x}, \mathbf{y}, \beta, \boldsymbol{\mu}) = \mathcal{L}_I(\mathbf{y}) + \beta\left(\sum_{i=1}^{K} x_i - 1\right) + \boldsymbol{\mu}^{\mathsf{T}}(A\mathbf{x} - \mathbf{y});$$

i.e., $\phi(\mathbf{x}, \mathbf{y}, \beta^*, \boldsymbol{\mu}^*) \le \phi(\mathbf{x}^*, \mathbf{y}^*, \beta^*, \boldsymbol{\mu}^*) \le \phi(\mathbf{x}^*, \mathbf{y}^*, \beta, \boldsymbol{\mu})$ *for all* $\mathbf{x}, \mathbf{y} \ge 0$, $\beta \in \mathcal{R}$, *and* $\boldsymbol{\mu} \in \mathcal{R}^T$.

Note that the second constraint in (4) reveals a general nonidentifiability property of the choice model.[4] Even if all possible offer sets are observed in the data (there are $O(2^n)$ of them), since there are potentially $O(n!)$ columns in the matrix A, there are more columns than rows, and the system of equations $A\mathbf{x} = \mathbf{y}$ is underdetermined; i.e., multiple solutions $\mathbf{x}$ could explain a given vector of aggregate likelihoods $\mathbf{y}$. In particular, given a sample of data (j_t, S_t), $t = 1, \ldots, T$, there could be multiple alternative optima $\mathbf{x}^*$ even though there is a unique λ^* (as explained above) and a unique $\mathbf{y}^*$, as shown by the following proposition.

Proposition 2. *If for each offer set* S, *and for each product* $j \in S$, *there is at least one observed transaction* (j, S) *in the data set, then there is a unique vector* $\mathbf{y}^*$ *that solves problem* (4).

Given our assumption that for the ground-truth choice model, $\mathbb{P}(j_t \mid S_t) > 0$ for all $j_t \in S_t$, the condition about the existence of an observation (j_t, S_t) for all $j_t \in S_t$ is mild. Proposition 2 implies that the choice model is *partially identifiable*, since it would be possible to learn the true values of λ and $\mathbb{P}(j_t \mid S_t)$ under an infinitely large sample. The identifiability of the choice model with respect to the arrival rate λ and the purchase probabilities $\mathbb{P}(j_t \mid S_t)$ makes the estimators λ^* and $\mathbf{y}^*$ inherit the desirable statistical properties of MLE estimators: they are consistent, asymptotically unbiased, and asymptotically efficient (i.e., attain the Cramér–Rao lower bound for the variance, asymptotically). The next result formalizes this observation.

Corollary 1. *Suppose the sequence of data grows such that for each set* S *observed in the sequence of offer sets,* $\lim_{T\to\infty}(\sum_{t=1}^{T} \mathbf{1}\{S_t = S\})/T = q_S > 0$, *a.s.; that is, each set observed appears infinitely often and appears in a positive fraction of the intervals, almost surely. Then as the sample size* $T \to \infty$, *the estimator* λ^* *satisfies* $\lambda^* \to \lambda$ *w.p.*1. *In addition, for each offer set* S *observed in the sequence,* $y_t^* \to \mathbb{P}(j_t \mid S_t)$ *w.p.*1, *and any optimal solution* $\mathbf{x}^*$ *to problem* (3) *satisfies* $A_t\mathbf{x}^* \to \mathbb{P}(j_t \mid S_t)$, *where* A_t *is the t-th row of* A.

[4] We say that the parameters $\mathbf{x}$ corresponding to the distribution over types $\boldsymbol{\sigma}$ is *identified* if for any alternative parameters $\mathbf{x}'$, $\mathbf{x} \neq \mathbf{x}'$, for some data (j_t, S_t), $t = 1, \ldots, \hat{T}$, we have $\mathcal{L}_I(\mathbf{x}) \neq \mathcal{L}_I(\mathbf{x}')$.

It is worth pausing here to point out some connections between our method and the robust optimization approach of Farias et al. (2013). The set of constraints in their formulation for revenue predictions (see Farias et al. 2013, §2.4) looks similar to our formulation (4). And whereas their matrix A that relates observed data and the underlying choice model is different,[5] our underlying definition of the feasible region shares a similar goal: to find a distribution of types compatible with the observed data. There are three major differences between the two approaches: First, Farias et al. seek to estimate worst-case revenues from a given assortment rather than finding MLEs for the pmf of types. Second, the approaches differ in how they address sampling noise. Farias et al. (2013, pp. 312–313 and §5.1.3) construct an uncertainty region $\mathcal{E}$ from the data. The uncertainty region $\mathcal{E}$ could be a box derived from sample averages of the associated choice probabilities and the corresponding confidence intervals. Then, they predict worst-case revenues from the region:

$$A\mathbf{x} = \mathbf{y}, \quad \mathbf{y} \in \mathcal{E}, \quad \text{and} \quad \mathbf{x} \ge 0, \quad \sum_{i=1}^{K} x_i = 1.$$

In our case, we do not impose any a priori constraint on $\mathbf{y}$, but the sample-based $\mathbf{y}^*$ that solves (4) satisfies the moment equations $A\mathbf{x}^* = \mathbf{y}^*$.

Finally, the computational approaches are also different. Farias et al. (2013) work on the dual formulation of the problem and, given the intractable number of constraints involved, solve it through constraint sampling, whereas we work on the primal problem and find directions of improvement via column generation, as described next.

4. Estimation Procedure

Given that we can observe arrivals/no arrivals and purchases/no purchases, our objective is to come up with a set of types $\boldsymbol{\sigma}$ that describe the data well and compute MLEs for the parameters $\mathbb{P}(\boldsymbol{\sigma})$. To this end, starting from an initial parsimonious set of types $\boldsymbol{\sigma}_0$, we develop a procedure to progressively "discover" new customer types that increase the likelihood value. This makes it possible to sequentially enrich the set of relevant types by building a sequence of sets $\{\boldsymbol{\sigma}_i\colon i = 1, 2, \ldots\}$.

4.1. Preprocessing

The market is assumed to be specified initially by a limited subset of customer types $\boldsymbol{\sigma}_0$. This initial set

[5] Farias et al. (2013) consider three different data representations based on proportions of pairwise comparisons for revealed preferences: comparison data, ranking data, and top-set data. See §§2.2 and 6.2 therein.

could be defined based on judgement and knowledge of the market or on marketing surveys—or could represent the *independent demand model*, described by a simple naïve set of n types who, respectively, rank only one of the products above the no-purchase alternative. Our method will iteratively augment this initial set.

We further assume that the set of customer types $\boldsymbol{\sigma}_0$ is *compatible* with the observed transactions and availability data, in the sense that the transactions observed can be explained by $\boldsymbol{\sigma}_0$. For instance, the independent demand model defined by singletons, spanning all possible products, guarantees such compatibility.[6]

4.2. Formulation of the Restricted Estimation Problem

Let $\boldsymbol{\sigma}$ be the current set of types under consideration (initially, $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0$), and let $\mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{x}, \lambda)$ be the incomplete data log-likelihood function (2) restricted to types $\boldsymbol{\sigma} = \{\sigma^{(1)}, \ldots, \sigma^{(N)}\}$. Assuming that the unique maximizer $\bar{\lambda}$ has already been established, the restricted MLE estimation problem can be formulated as follows:

$$\begin{aligned} \max_{\mathbf{x}\geq 0} \quad & \mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{x}) \\ \text{s.t.} \quad & \sum_{i=1}^{N} x_i = 1. \end{aligned} \tag{6}$$

Formulation (6) is a concave, constrained optimization problem for a given set of customer types $\boldsymbol{\sigma}$, defined over the open set $\mathbf{x} \geq 0$, with a global (though not necessarily unique) optimum $\bar{x}_1, \ldots, \bar{x}_N$. It can be solved using standard nonlinear optimization methods or via an efficient EM method that we developed in a related paper (van Ryzin and Vulcano 2013).

For our illustrative example described in Table 1, problem (6) becomes

$$\begin{aligned} \max_{x_1, x_2 \geq 0} \quad & \{4\log(x_1 + x_2) + 3\log(x_2) + \log(x_1)\} \\ \text{s.t.} \quad & x_1 + x_2 = 1, \end{aligned}$$

leading to MLE estimates $\bar{x}_1 = 0.25$ and $\bar{x}_2 = 0.75$. The estimated arrival rate would be $\bar{\lambda} = 0.8$.

4.3. Discovering New Types

We next examine the problem of augmenting a given set of customer types $\boldsymbol{\sigma}$ with other types from the factorial number of candidates.

First, note that from the optimal solution to problem (6), we can easily construct a feasible solution for the estimation problem (3) over the full set of types, with $x_1 = \bar{x}_1, \ldots, x_N = \bar{x}_N$, and $x_{N+1}, \ldots, x_K = 0$.

For types $\boldsymbol{\sigma}$, define the matrix $A \in \{0, 1\}^{\hat{T} \times N}$, with elements $a_{ti} = 1$ if $i \in \mathcal{M}_t(j_t, S_t)$, and $a_{ti} = 0$ otherwise. Consider the following problem restricted to the set of types $\boldsymbol{\sigma}$:

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{y}\geq 0} \quad & \mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{y}) \\ \text{s.t.} \quad & \sum_{i=1}^{N} x_i - 1 = 0, \\ & A\mathbf{x} - \mathbf{y} = 0. \end{aligned} \tag{7}$$

This problem is also a concave program with linear constraints and bounded objective $\mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{y}) \leq 0$ for all $\mathbf{x}, \mathbf{y} \geq 0$. Take the maximum $\bar{\mathbf{x}}$ of problem (6). Note that the point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, where $\bar{y}_t = \sum_{i\in\mathcal{M}_t(j_t, S_t)} \bar{x}_i$, is an optimal solution to problem (7).

The optimality conditions previously presented in Proposition 1 lead to a sequential procedure to discover new types. Note that the estimation problem (7) restricted to types $\boldsymbol{\sigma}$ is also a concave program for which these optimality conditions hold. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\beta}, \bar{\boldsymbol{\mu}})$ be a saddle point of the associated (restricted) Lagrangian $\phi^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \beta, \boldsymbol{\mu})$, with $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ being an optimal (primal) solution to problem (7), and where $\bar{y}_t = \sum_{i\in\mathcal{M}_t(j_t, S_t)} \bar{x}_i$, for $t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda$.

Define $G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}) = \phi^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \bar{\beta}, \bar{\boldsymbol{\mu}})$. From Theorem 6.2.6 in Bazaraa et al. (2006), $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\beta}, \bar{\boldsymbol{\mu}})$ is also a Karush-Kuhn-Tucker (KKT) point and has to satisfy the following conditions:

- For $t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda$,

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y})}{\partial y_t}\bigg|_{(\mathbf{x}, \mathbf{y})=(\bar{\mathbf{x}}, \bar{\mathbf{y}})} = \frac{1}{\bar{y}_t} - \bar{\mu}_t = 0, \quad \text{and then } \bar{\mu}_t = \frac{1}{\bar{y}_t}. \tag{8}$$

Note that $\bar{u}_t > 1$.

- The value $\bar{\beta}$ can be computed from the derivative with respect to any x_i:

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y})}{\partial x_i}\bigg|_{(\mathbf{x}, \mathbf{y})=(\bar{\mathbf{x}}, \bar{\mathbf{y}})} = \bar{\beta} + \sum_{t:\, i\in\mathcal{M}_t(j_t, S_t)} \bar{\mu}_t = 0 \quad \text{for all } i. \tag{9}$$

Given $A\bar{\mathbf{x}} = \bar{\mathbf{y}}$, observe that $(\bar{\boldsymbol{\mu}}^\top A)\bar{\mathbf{x}} = \bar{\boldsymbol{\mu}}^\top \bar{\mathbf{y}} = \hat{T}$, since from (8), $\bar{\mu}_t \bar{y}_t = 1$ for $t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda$. Thus,

$$\begin{aligned} (\bar{\boldsymbol{\mu}}^\top A)\bar{\mathbf{x}} &= \sum_{i=1}^{N}\bigg(\sum_{t:\, i\in\mathcal{M}_t(j_t, S_t)} \bar{\mu}_t\bigg)\bar{x}_i = \sum_{i=1}^{N}(-\bar{\beta})\bar{x}_i \quad \text{(from (9))} \\ &= (-\bar{\beta})\sum_{i=1}^{N}\bar{x}_i = \hat{T}, \end{aligned}$$

so that $\bar{\beta} = -\hat{T}$.

[6] For the illustrative example, from Table 1, it is enough to define $\boldsymbol{\sigma}_0 = \{(1), \ldots, (5)\}$ to have a *compatible* set of types. On the other hand, the set of types $\boldsymbol{\sigma}_0 = \{(2, 3), (3, 4, 5)\}$ is incompatible, since it cannot explain the purchase of product 1.

Next, consider again the full set of types and the associated dual function

$$\Theta(\bar{\beta}, \bar{\boldsymbol{\mu}}) = \sup_{\mathbf{x}, \mathbf{y} \geq 0} \{\phi(\mathbf{x}, \mathbf{y}, \bar{\beta}, \bar{\boldsymbol{\mu}})\}.$$

For the unrestricted problem, define $G(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}, \mathbf{y}, \bar{\beta}, \bar{\boldsymbol{\mu}})$. If for all i, $N+1 \leq i \leq K$, $\partial G(\mathbf{x}, \mathbf{y})/\partial x_i \leq 0$, then $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\beta}, \bar{\boldsymbol{\mu}})$ is a saddle point of the Lagrangian, and from Proposition 1, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is an optimal solution to primal problem (4) (and hence to original problem (3)). Otherwise, there exists k, $N+1 \leq k \leq K$, such that

$$\frac{\partial G(\mathbf{x}, \mathbf{y})}{\partial x_k} = \sum_{\substack{t:\ \sigma^{(k)}(j_t) < \sigma^{(k)}(i) \\ \forall i \in S_t,\, i \neq j_t}} \bar{\mu}_t > \hat{T}, \tag{10}$$

and customer type k defines a direction of improvement for the Lagrangian dual function.

Condition (10) has an intuitive interpretation. Consider the new type k and a perturbation of the solution $\bar{\mathbf{x}}$, denoted by $\bar{\mathbf{x}}(\epsilon)$, that shifts ϵ probability mass to type k and reduces the probability mass of the current types by a factor $1-\epsilon$ so that total probability is preserved: $x_k(\epsilon) + \sum_{i=1}^{N} \bar{x}_i(\epsilon) = \epsilon + \sum_{i=1}^{N} (1-\epsilon)\bar{x}_i = 1$. To analyze this perturbation, note that the dual variable $\bar{\mu}_t$ gives the marginal change in log-likelihood with respect to changes in the probability $\bar{y}_t$ of the outcome observed in period t (i.e., $\partial \mathcal{L}_I(\mathbf{y})/\partial y_t = 1/y_t = u_t$). Reducing the probability of the current types by a multiplicative factor $(1-\epsilon)$ reduces the sum of any combination of x_i's (i.e., $\bar{y}_t$) by the same factor. Taking $\mathcal{L}_I(\mathbf{y})$ and Taylor-expanding it around $\bar{y}_t$, we observe that the log-likelihood decreases by $\epsilon \sum_{t=1}^{\hat{T}} \bar{\mu}_t \bar{y}_t = \epsilon \hat{T}$, since from (8), $\bar{\mu}_t \bar{y}_t = 1$ for all t. On the other hand, the new type k adds probability mass ϵ to each period in which type k is compatible with the observed outcome and hence increases the log-likelihood by $\epsilon \sum_{t:\ \sigma^{(k)}(j_t) < \sigma^{(k)}(i), \forall i \in S_t,\, i \neq j_t} \bar{\mu}_t$. Condition (10) simply says that if the current set of types is not optimal, then there must be a new type k for which the total effect of this perturbation is positive, i.e., increases the likelihood.

This observation means we can solve the primal problem over a limited yet augmented set of types $\boldsymbol{\sigma}' = \boldsymbol{\sigma} \cup \{\sigma^{(k)}\}$, and we get a new primal solution $(\bar{\bar{\mathbf{x}}}, \bar{\bar{\mathbf{y}}})$ so that

$$\mathcal{L}_I^{\boldsymbol{\sigma}'}(\bar{\bar{\mathbf{y}}}) \geq \mathcal{L}_I^{\boldsymbol{\sigma}}(\bar{\mathbf{y}}),$$

and hence

$$\phi^{\boldsymbol{\sigma}'}(\bar{\bar{\mathbf{x}}}, \bar{\bar{\mathbf{y}}}, \bar{\bar{\beta}}, \bar{\bar{\boldsymbol{\mu}}}) \geq \phi^{\boldsymbol{\sigma}}(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\beta}, \bar{\boldsymbol{\mu}}),$$

where $\bar{\bar{\beta}} = \bar{\beta} = -\hat{T}$, and $\bar{\bar{\boldsymbol{\mu}}}$ are dual variables associated with the new primal problem based on types $\boldsymbol{\sigma}'$. By repeating this argument, since the number of potential types is finite and at any step we either add a new type or reach optimality, the procedure is guaranteed to converge to a global optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$.

4.3.1. Complexity of the Type Discovery Subproblem. Generating an improving customer type $(N+1)$ can be posed as an optimization problem: given sets S_t and coefficients $\bar{\mu}_t$, $t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda$, we are interested in maximizing the sum in (10). This problem is related to a single-machine scheduling problem with $(n+1)$ jobs of unit length (recall that there are n products plus the no-purchase option) known as the *linear ordering problem* (e.g., see Martí and Reinelt 2011). An important difference between our formulation and the standard linear ordering problem in the literature is that in our case, the cost structure is not defined at the job–pair level (i.e., there is no explicit cost for having job j_1 before job j_2), but rather it is defined at the bundle level (i.e., there is a cost for having a job scheduled earlier than a set of other jobs). In addition, here we are maximizing benefits rather than minimizing cost.

The linear ordering problem is known to be NP-hard in its general form. Still, given the particular structure of the variation that we are considering here, its complexity is worth verifying formally. The following proposition establishes that the NP-hard, maximum independent set problem (see Garey and Johnson 1979) can be reduced to (10), so that our market discovery problem is also NP-hard. In fact, given the reduction from maximum independent set, our problem is strongly NP-hard; i.e., there is not even a fully polynomial-time approximation scheme unless P = NP (see Zuckerman 2007).

Proposition 3. *The following optimization problem related to our market discovery algorithm is NP-hard*:

$$\begin{aligned} &\max \left\{ \sum_{t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda} \bar{\mu}_t w_t \right\} \\ &\text{s.t.}\quad w_t \leq \mathbf{I}\{\sigma(j_t) < \sigma(i)\ \forall i \in S_t,\, i \neq j_t\}, \\ &\qquad\qquad t \in \mathcal{P} \cup \bar{\mathcal{P}}_\lambda. \end{aligned} \tag{11}$$

4.3.2. A Mixed Integer Programming Formulation to Discover New Types. Finding an improving customer type reduces to solving (11) and checking whether the solution satisfies condition (10). The problem can be formulated as a mixed integer program (MIP).

Define binary linear ordering variables x_{ji}, which are equal to 1 if product j goes before i in the sequence (i.e., preference list) and equal to 0 otherwise. In addition, define binary variables w_t, which are equal to 1 if the new type should be added to the set $\mathcal{M}_t(j_t, S_t)$ and equal to 0 otherwise. Overall, there are $O(n^2+\hat{T})$ binary variables and $O(n^3+nT)$ constraints. The MIP is as follows:

$$\begin{aligned}
\max \quad & \sum_{t\in\mathcal{P}\cup\bar{\mathcal{P}}_\lambda} \bar{\mu}_t w_t \\
\text{s.t.}\quad & x_{ji}+x_{ij}=1 \quad \forall j,i,\ 0\le j<i\le n,\\
& x_{ji}+x_{il}+x_{lj}\le 2\\
& \qquad\qquad \forall j,i,l,\ j\ne i\ne l,\ 0\le j,i,l\le n,\\
& \sum_{j=1}^{n} x_{j0}\ge 1,\\
& w_t\le x_{j_t,i} \quad \forall i\in S_t,\ j_t\ne i,\ \text{and } t\in\mathcal{P}\cup\bar{\mathcal{P}}_\lambda,\\
& x_{ji},w_t\in\{0,1\},\quad 0\le j,i\le n,\ t\in\mathcal{P}\cup\bar{\mathcal{P}}_\lambda.
\end{aligned} \tag{12}$$

The first set of equalities ensures that either product j goes before product i in the preference list, or product i goes before product j. The second set represents transitivity constraints that ensure a linear ordering among three products. The third constraint precludes the no-purchase alternative from being the most preferred option. The fourth set of constraints guarantees that in period t, product j_t must be preferred over all other available products in S_t for the proposed type to make a contribution to the objective function in period t. Note that $w_t=1$ only if $x_{j_t,i}=1$, for all $i\in S_t$, $i\ne j_t$.

If the optimal objective function value is greater than $\hat{T}$, then we have found a customer type that will increase the value $\mathcal{L}_I(\mathbf{x})$ in (3). The position of product j in the new ranking, $\sigma^{(N+1)}(j)$, can be determined from $\sigma^{(N+1)}(j)=\sum_{i\in\mathcal{N},\,i\ne j} x_{ij}+1$.

4.3.3. Implementation. The discovery mechanism starts from an initial set $\boldsymbol{\sigma}_0$ of *compatible* customer types. This initial set can be built on an informed guess of the market composition. Alternatively, barring any prior knowledge on the market, we can start with a naïve independent demand model where we have a customer type $\sigma^{(j)}=(j)$ for each product j, leading to n singleton types.

Then, we start a sequence of estimation–discovery stages. If $\boldsymbol{\sigma}$ is the current set of N types under consideration (initially, $\boldsymbol{\sigma}=\boldsymbol{\sigma}_0$), we solve the estimation problem (6) and get proportions $\bar{x}_i$, $1\le i\le N$, and dual variables $\bar{\mu}_t$. Next, we apply our market discovery algorithm using the mixed integer programming formulation (12). We point out that does not need to be solved to optimality; it is enough to find a solution that satisfies (10) (or a certificate that such solution does not exist). Even though this mechanism provides a customer type that defines a direction of improvement for the log-likelihood function when such a direction does exist, we perform a log-likelihood ratio test to check the statistical significance of the new type. Note that the set of types of iteration i, $\boldsymbol{\sigma}_i$, is nested in the set of types of iteration $i+1$, $\boldsymbol{\sigma}_{i+1}$.[7] For a given entering type $N+1$, we set a null hypothesis H_0: $x_{N+1}=0$. Consider the log-likelihood values $\mathcal{L}_I^{\boldsymbol{\sigma}_{i+1}}(\bar{\mathbf{x}},x_{N+1})$, with $x_{N+1}\ge 0$, and $\mathcal{L}_I^{\boldsymbol{\sigma}_i}(\bar{\mathbf{x}},0)$. The large sample distribution of $-2(\mathcal{L}_I^{\boldsymbol{\sigma}_i}(\bar{\mathbf{x}},0)-\mathcal{L}_I^{\boldsymbol{\sigma}_{i+1}}(\bar{\mathbf{x}},x_{N+1}))$ is chi-squared, with one degree of freedom (e.g., see Greene 2003, Theorem 17.5). The (low) critical value for this statistic is 3.842 at the 5% significance level. This sequence of estimation–discovery stages is repeated until a convergence criteria is met: either a maximum number of new types is reached or the likelihood ratio test is not passed. It is worth pointing out that the set of constraints of (12) does not change from iteration to iteration; only the coefficients $\bar{\mu}_t$ in the objective function are updated.

Note that this implementation does not explicitly account for model complexity and the potential of overfitting when building up the set of types. In our numerical experiments (see §5), we apply the procedure as described using only the likelihood ratio stopping criterion with no explicit complexity penalties. Nevertheless, the resulting models achieve a very good out-of-sample performance, even when adjusted for complexity as measured by the corrected Akaike information criterion (AIC$_c$, defined in §5). Moreover, it is not difficult to adjust the procedure to explicitly account for model complexity. For example, one could use the AIC$_c$ as an alternative stopping criterion, terminating the process of generating new types when the AIC stops decreasing. Alternatively, one could append an ad hoc complexity penalty on the number of types to the likelihood objective and again terminate the procedure when the penalized likelihood stops increasing. Again, however, our numerical experience with the procedure suggests that the likelihood function flattens out rapidly after adding only a small number of types, and hence the procedure terminates well before overfitting becomes an issue.

5. Numerical Examples

We next present the results of an extensive set of tests of our market discovery algorithm on both simulated and real-world data sets. We implemented it using the MATLAB[8] procedural language and its interface with IBM ILOG CPLEX v.12.4 to solve the associated MIP. The log-likelihood function was optimized using the EM method detailed in van Ryzin and Vulcano (2013). In the following examples, we made the simplification that $\lambda=1$ (i.e., there is exactly one arrival per period).[9]

[7] One model is considered nested in another if the first model can be generated by imposing restrictions on the parameters of the second.

[8] MATLAB is a trademark of The MathWorks, Inc. We used version 8.0 (R2012b) for OS X on a Mac with a 2.6 GHz Intel Core i7 processor and 8 GB of RAM.

[9] Recall that for the general case, the estimate for λ can be easily computed via the formula $\bar{\lambda}=\hat{T}/T$, with $\hat{T}=|\mathcal{P}|+|\bar{\mathcal{P}}_\lambda|$.

5.1. Experiments Based on Synthetic Data

The goal of our experiments based on synthetic data is to understand how accurately our market discovery algorithm identifies a known, underlying choice structure. For this first set of experiments, we focus on three parametric models used extensively in choice modeling practice: the MNL, cross-nested logit (CNL), and latent class multinomial logit (LC-MNL) models. First, for a given set of n products, we generate availability data over T periods. Then, starting from a parametric characterization of each of the models and given the availability data, we generate transaction data, which resembles the data one would have access to in practice. We use this transaction data as a training data set and run our market discovery algorithm starting from n singleton customer types (i.e., the independent demand assumption). The output of the procedure is an enriched set of customer types. Next, we generate a new availability data set and use our discovered set of customer types to estimate, for each period t, the likelihood that a transaction for each available product $j \in S_t$ occurs (via Equation (1)): $\hat{\mathbb{P}}(j \mid S_t) = \sum_{i \in \mathcal{M}_t(j, S_t)} \bar{x}_i$ for $j \in S_t$. We then compare these estimates to the *true* likelihoods computed using the known ground-truth model.

We calculate two measures of goodness of fit to compare the performance of both the independent demand assumption and our market discovery (MD)-based customer types versus the ground-truth probabilities: root mean square error (RMSE) and AIC_c. The RMSE is an absolute measure of fit between estimated probabilities $\hat{p}_q$ and true probabilities p_q, defined as $\mathrm{RMSE} = (\sum_{q=1}^{Q} (\hat{p}_q - p_q)^2)/Q)^{1/2}$, where Q is the number of observations evaluated. A value of zero indicates perfect fit. One downside is that RMSE does not include any penalty for model complexity, which increases as we add types to the original singletons. To account for this, we also report the corrected Akaike information criterion, defined as $\mathrm{AIC}_c = 2(N - \mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{x}) + N(N+1)/(T-N-1))$, where N is the number of parameters (i.e., types) in the model, $\mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{x})$ is the maximized value of the log-likelihood function for a given set of types $\boldsymbol{\sigma}$, and T is the sample size (i.e., number of periods). The AIC_c measure rewards the log-likelihood value but also penalizes model complexity (captured by the number of parameters) to control for overfitting. Lower values of AIC_c indicate a better fit.

Recall that under the current uncensored demand setting, the modeler observes all arrivals and therefore has full information about the market share (i.e., the ratio between number of transactions and number of arrivals).

5.1.1. The Amazon Model: Setup and Variations.

Following the data description in Farias et al. (2013, §4.2.1), we consider an MNL model fit to DVD sales data from Amazon.com collected during three months in 2005. The experiments are based on $n = 15$ substitutable DVDs, where the customer's utility for a given DVD j has the form $U_j = u_j + \xi_j$, where u_j is the nominal utility and ξ_j is a standard Gumbel random variable. The *weight* of product j is given by $w_j = \exp(u_j)$. The weights as reported by Farias et al. are given in the online appendix (see Table A1, segment #1).

A summary of the structure of each of the models used to generate the synthetic data is included below. For each of the models, we generate a training data set containing $T = 500$ random offer sets of between one and seven products and 500 corresponding transactions and no purchases simulated based on the given structural model. Then, we generate another 1,000 random assortments of between one and seven items to assess the out-of-sample performance of our market discovery algorithm. Train (2003) provides an excellent overview of the three models summarized below.

The MNL model: For this model, the probability that product j is chosen in period t when assortment S_t is offered is given by

$$\mathbb{P}(j \mid S_t) = \frac{w_j}{\sum_{h \in S_t} w_h}, \tag{13}$$

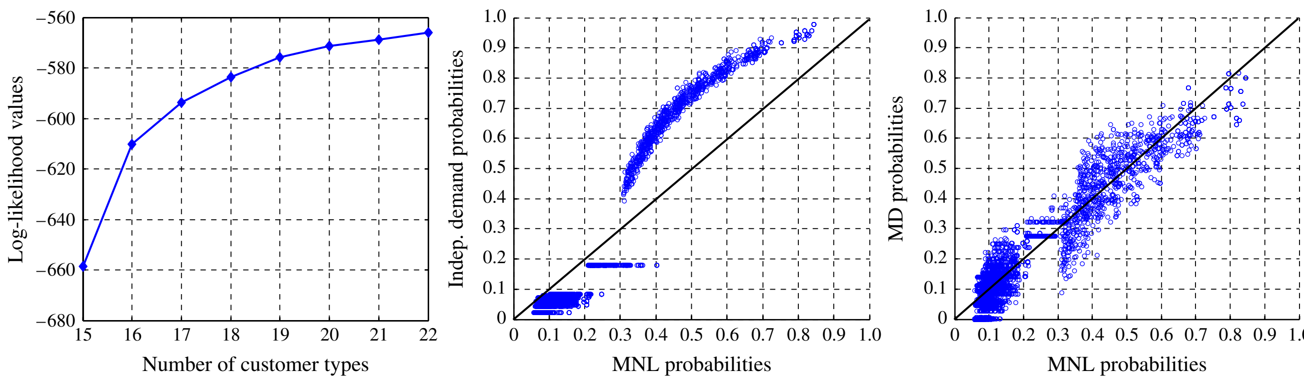
with $0 \in S_t$. Assuming a market share of 80% for this category (i.e., a no-purchase probability of 20%), the utility of the no-purchase option is determined to be $u_0 = -3.197$, with weight $w_0 = 0.04085$.

Figure 1 illustrates the performance of our MD algorithm. In the left panel we plot the evolution of the log-likelihood function value, starting from the independent demand assumption with $N = 15$ singleton types. The first observation in the horizontal axis is the result of maximizing the log-likelihood function, yet constrained to the original 15 singleton types. The last seven observations reflect the behavior of our MD algorithm when sequentially adding significant types. The algorithm was terminated when the next type found did not provide a statistically significant increment in the log-likelihood function.

The scatterplot in the center panel depicts the estimated independent demand probabilities versus the ground-truth MNL probabilities, where the 45° line represents perfect fit. That is, we plot pairs $(\mathbb{P}_{\mathrm{MNL}}(j_t \mid S_t), \hat{\mathbb{P}}_{\mathrm{Indep}}(j_t \mid S_t))$, $j_t \in S_t$ (possibly, $j_t = 0$) over the 1,000 periods in the out-of-sample data set.

The scatterplot in the right panel depicts the predicted nonparametric-based probabilities—computed after running the MD algorithm and adding seven customer types to the set—versus the ground-truth MNL probabilities.

Figure 1 (Color online) Performance of the MD Algorithm When Facing an Underlying MNL Model



Notes. Left, evolution of the log-likelihood function within MD. Center, predicted nonzero probabilities under the independent demand assumption versus ground-truth probabilities. Right, predicted nonzero probabilities under MD versus ground-truth probabilities.

The CNL model: For this model, the set of products is partitioned into L subsets, or "nests," denoted by $\mathcal{N}_1, \ldots, \mathcal{N}_L$. The probabilities take the form

$$\mathbb{P}(j \mid S_t) = \mathbb{P}(\mathcal{N}_l \mid S_t)\mathbb{P}(j \mid \mathcal{N}_l, S_t) = \frac{w(l, S_t)^\rho}{\sum_{h=1}^{L} w(h, S_t)^\rho} \frac{w_j}{w(l, S_t)},$$

where $\rho < 1$ is a scale parameter, and

$$w(l, S_t) = \alpha_l w_0 + \sum_{i \in (\mathcal{N}_l \cap S_t) \setminus \{0\}} w_i.$$

Like in the MNL case, we assume $w_0 = 0.04085$.

The parameter α_l captures the level of membership of the no-purchase option in nest l and satisfies $\sum_{l=1}^{L} \alpha_l^\rho = 1$, $\alpha_l \geq 0$, $l = 1, \ldots, L$. The CNL is a variation of the standard nested logic model where a particular alternative can belong simultaneously to multiple nests. Like in Farias et al. (2013), we assume that the no-purchase option belongs to all nests and all other products have membership in only one nest. We also partition the products into four nests with the first nest containing products 1–5, the second nest containing products 6–9, the third nest containing products 10–13, and the last nest containing products 14 and 15. We chose $\rho = 0.5$ and assign the no-purchase option to every nest with nest membership parameter $\alpha_l = (1/4)^{1/\rho} = 1/16$.

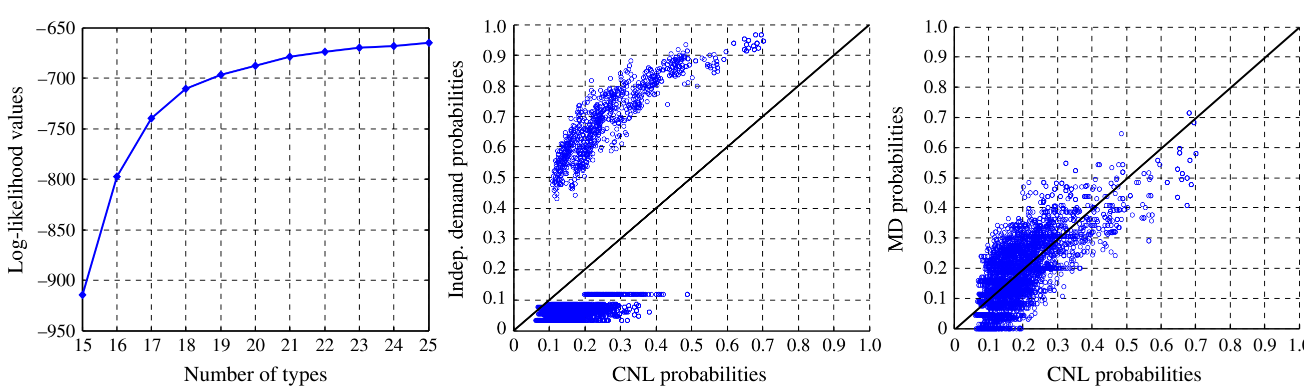
Figure 2 illustrates the performance of our MD algorithm when executed on data generated by a ground-truth CNL model, in the same manner as we do for the MNL model.

The LC-MNL model: The LC-MNL is a flexible model that can be interpreted as a weighted average MNL, with C classes, class-dependent weights, and a market composition described by proportions $p_1, \ldots, p_C$ that capture the heterogeneity in taste. The probability that product j is chosen in period t when assortment S_t is offered is given by

$$\mathbb{P}(j \mid S_t) = \sum_{c=1}^{C} p_c \frac{w_{j,c}}{\sum_{h \in S_t} w_{h,c}},$$

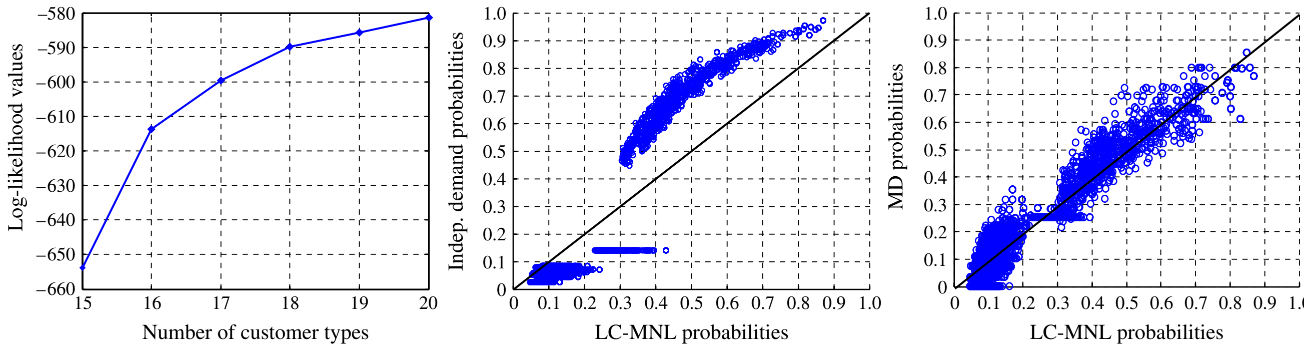
where $w_{j,c}$ is the weight that a customer of segment c gives to product j. The proportions satisfy $p_c \geq 0$,

Figure 2 (Color online) Performance of the MD Algorithm When Facing an Underlying CNL Model



Notes. Left, evolution of the log-likelihood function within MD. Center, predicted nonzero probabilities under the independent demand assumption versus ground-truth probabilities. Right, predicted nonzero probabilities under MD versus ground-truth probabilities.

Figure 3 (Color online) Performance of the MD Algorithm When Facing an Underlying LC-MNL Model



Notes. Left, evolution of the log-likelihood function within MD. Center, predicted nonzero probabilities under the independent demand assumption versus ground-truth probabilities. Right, predicted nonzero probabilities under MD versus ground-truth probabilities.

$\sum_{c=1}^{C} p_c = 1$. We consider a variation of the basic Amazon MNL model with $C = 5$ segments, weights, and proportions of the different segments described in Table A1 in the online appendix. For each segment, the no-purchase preference $w_{0,c}$ was adjusted to give a market share of 80% within the segment.

Figure 3 illustrates the performance of our MD algorithm when executed on data generated by an LC-MNL model, in the same manner as we do for both previous models.

Discussion of results: The behavioral pattern of our algorithm is similar over the three structural models. As expected, in the left panel of Figures 1–3, we observe a decreasing marginal benefit of adding new types, which leads the MD algorithm to stop after a handful of iterations.

The three center panels show the performance of the independent demand model. Most of the points (around 4,000 in each of the charts) are located in the lower left corner and correspond to estimates of the probability of purchasing a product. Recall that this is a case of high market share, and therefore each customer is very likely to substitute in case of stock-out of her most preferred product under any of the choice-based demand models. The true probabilities of purchasing a given product are below 0.5 across the three models. The independent demand model tends to underestimate these probabilities, since an arrival that does not find her single choice available leads to a no purchase, which in turn overestimates the no-purchase probability, as reflected by the (approximately 1,000) points located above the 45° line.

The panels on the right depict the predicted nonparametric-based probabilities versus the true choice-based demand probabilities. The points are better aligned around the 45° line, indicating the correction introduced by the MD algorithm when starting from the simple independent demand model.

The number of periods in the training data set plays a fundamental role in the quality of the MD estimates, independently of the true demand model, as shown by Table 2. Fixing the number of periods in the out-of-sample data set at 1,000, we tested different sizes of training data sets (Figures 1–3 correspond

Table 2 Performance of the Independent Demand Assumption and the MD Algorithm Under Different Sizes of Training Data Sets, T

		Indep. demand performance			MD performance			
Ground-truth	T	RMSE	LL value	AIC$_c$	No. of types	RMSE	LL value	AIC$_c$
MNL	100	0.11	−147.28	330.27	18	0.11	−108.17	260.78
	300	0.11	−424.22	880.13	20	0.08	−352.94	748.89
	500	0.11	−660.40	1,351.79	20	0.06	−581.94	1,205.63
	700	0.11	−993.39	2,017.48	22	0.06	−865.98	1,777.45
CNL	100	0.22	−183.32	402.35	20	0.12	−113.31	277.26
	300	0.21	−508.26	1,048.21	22	0.08	−370.65	788.95
	500	0.21	−915.04	1,861.07	25	0.08	−664.98	1, 382.70
	700	0.21	−1,301.20	2,633.10	22	0.08	−972.73	1,990.95
LC-MNL	100	0.11	−134.91	305.53	17	0.09	−105.47	252.40
	300	0.11	−423.73	879.15	19	0.07	−357.40	755.51
	500	0.11	−653.84	1,338.67	20	0.05	−581.37	1,204.49
	700	0.11	−959.92	1,950.54	22	0.05	−818.31	1,682.11

Note. The initial number of types is $N = 15$, and the market share is 0.8. LL, log-likelihood.

to $T = 500$). As expected, the quality of the estimates measured by RMSE tends to improve as the number of periods in the training set increases. The number of types added by the MD algorithm also tends to increase as the number of periods grows. Yet the lower AIC_c value for MD-based estimates compared with the independent demand estimates indicates that the additional predictive power of our algorithm does not come at the expense of excessive model complexity and does not suffer from significant overfitting.

The computational burden also grows significantly as we increase T; for the case $T = 100$, it took approximately 0.1 seconds to solve the MIP and find a new customer type under the MNL model, whereas the time to find a new type when $T = 500$ averaged 16 seconds.

Maintaining the number of in-sample periods at $T = 500$, we analyzed the effect of different market shares by adjusting the utility of the no-purchase option (see Table A2 in the online appendix). When the share is low to moderate (i.e., below 50%), we observe no purchases in most of the periods, and the independent demand assumption becomes a good predictor of the consumers' choice behavior, with no need to add new types. This is because when a customer arrives and her first choice is not available, it is very likely that she does not purchase given the high no-purchase utility.

Based on our experiments, we observe that for the Amazon model under consideration (15 products, between one and seven products available per period), the MD algorithm needs at least 100 to 150 observed purchases to gather enough information and improve the predictions of the basic independent demand model.

5.1.2. The Amazon Model: Sensitivity Analysis. So far we have tested three structurally diverse models, each for one set of parameter values. In this section, we perturb the original parameters of these examples to generate more instances of data to check the robustness of the MD algorithm. For each structural family, we generate 50 perturbations of the model parameters, including randomly generated market shares between 50% and 80%. Then, for each perturbation, we ran the MD algorithm on a training data set of 500 periods, where in each period there are one to seven randomly selected products available, as we do for the original case. We then tested the performance of the estimated model (i.e., set of customer types and associated pmf) on an out-of-sample data set consisting again of one to seven available products per period over 1,000 periods. Next, we computed the MD-based estimate of purchase probabilities in a given out-of-sample period, the corresponding ground-truth probabilities, and the relative error between both. That is, if $\mathbb{P}(j \mid S_t)$ is the true probability of purchasing $j = 0, 1, \ldots, 15$, for $j \in S_t$, and $\hat{p}_{j,t}$ is its MD-based estimate, we computed the relative error $(\hat{p}_{j,t} - \mathbb{P}(j \mid S_t))/\mathbb{P}(j \mid S_t)$. To assess the impact of structural errors from that of sampling errors, we also computed maximum likelihood estimates assuming knowledge of the structure of the underlying demand model. We computed these structural estimates on the same training data set used for the MD estimates and then also computed the relative errors with respect to the true probabilities based on the same out-of-sample data set.

According to Table A1 (segment #1) in the online appendix, the original, nominal utilities for the 15 products vary between −4.89 and −3.59. For the perturbed MNL instances, we drew nominal utilities for all the products from a Uniform(−5, −3) distribution. To get maximum likelihood MNL estimates, we maximize the function

$$\mathscr{L}_{\mathrm{MNL}}(\mathbf{w}) = \sum_{t=1}^{T} \log(\mathbb{P}(j_t \mid S_t))$$

$$\text{subject to } \frac{\sum_{j=1}^{15} w_j}{\sum_{j=1}^{15} w_j + w_0} = \alpha,$$

where $\mathbb{P}(j_t \mid S_t)$ is given by (13) and the market share α is generated randomly in the range (0.5, 0.8). Since there is a continuum of maxima, for each perturbed instance we set w_0 equal to the true underlying w_0.

Figure 4 represents distributions of the relative error for the MNL estimates (left) and MD estimates (right). The quartiles for the MNL errors are $(-0.157, -0.008, 0.123)$, whereas for the MD errors, they are $(-0.187, 0.032, 0.431)$, indicating a small underestimation bias of the MNL estimates and an overestimation bias of the MD estimates. The vertical bar at error (−1) for the MD errors stands for the probabilities that are (almost) zero under the MD estimated model but nonzero and small under the true model. They account for approximately 7.5% of the total sample size.

Likewise, for each perturbed CNL instance, we maintain the nests and values $\rho = 0.5$ and $\alpha_l = 1/16$ and drew product nominal utilities from a Uniform(−5, −3) distribution. We also generated market shares in the range (0.5, 0.8) and adjusted w_0 accordingly. Then, we applied maximum likelihood estimation by assuming that the modeler has complete information about ρ and α_l, which is a favorable setting for CNL. Figure 5 shows the resulting distributions of the relative error for CNL estimates (left) and MD estimates (right). We observe that even for the favorable case of having perfect information

Figure 4 (Color online) Relative Error of Probabilities Estimated by Maximum Likelihood Based on MNL (Left) and the MD Algorithm (Right) with Respect to 50 Perturbed, Ground-Truth MNL Models

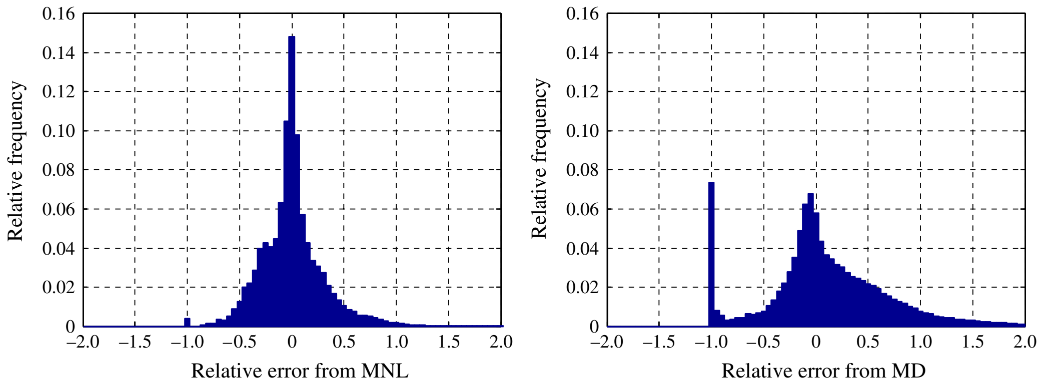
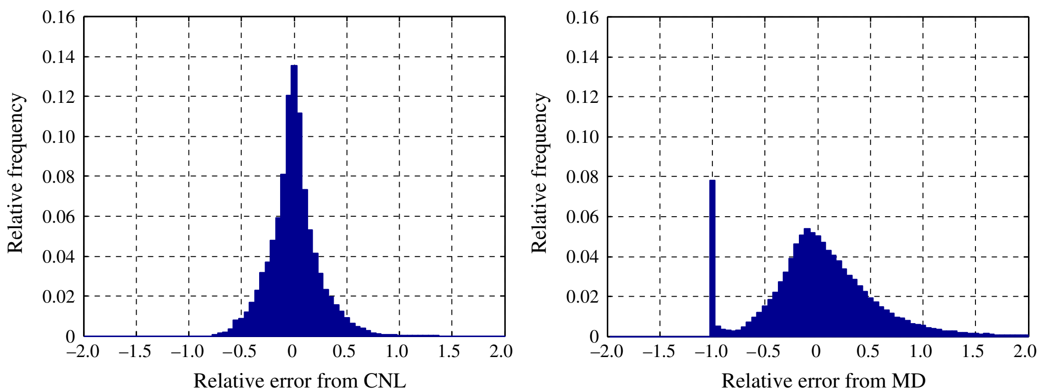


Figure 5 (Color online) Relative Error of Probabilities Estimated by Maximum Likelihood Based on CNL (Left) and the MD Algorithm (Right) with Respect to 50 Perturbed, Ground-Truth CNL Models

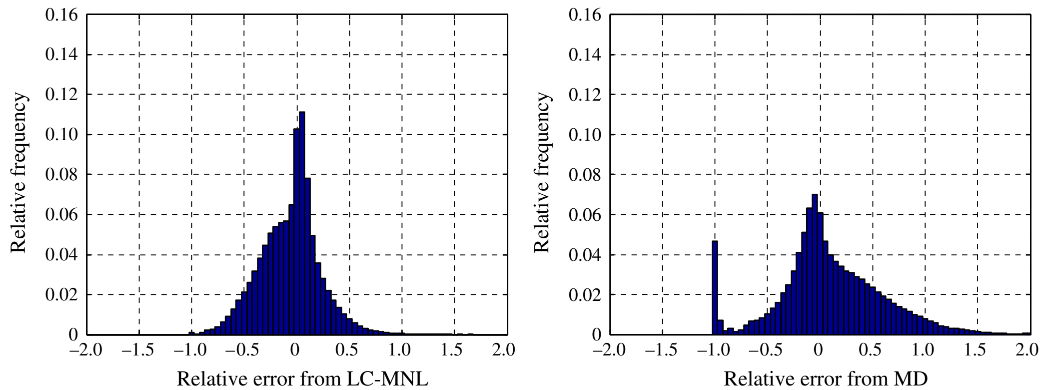


about the choice-based demand structure and knowing a few parameters of the true model, the CNL estimation error is significant. This is due to the higher complexity of this model compared with the previous MNL. The quartiles for the CNL errors are $(-0.125, -0.007, 0.107)$, and for the MD errors, they are $(-0.266, -0.010, 0.295)$, indicating quite centered estimates for both cases and a wider distribution for the MD errors.

Finally, for the perturbed LC-MNL instances, we assume that there are five segments in the market, with randomly generated proportions. Then, for each segment, we sampled product nominal utilities from a Uniform$(-5, -3)$ distribution. We also generated market shares in the range (0.5, 0.8), and for each segment c, we adjusted $w_{0,c}$ accordingly. Then, we applied maximum likelihood estimation by assuming that the modeler has complete information about the number of segments in the market and of the even split of the no-purchase option across the different segments. Figure 6 shows the resulting distributions of the relative error for the LC-MNL estimates (left) and MD estimates (right). We observe that for the favorable case of having perfect information about the choice-based demand structure and even perfect information about a few parameters of the true model, the estimation error is significant. This is due to the higher complexity of this model compared with the previous CNL. The quartiles for the LC-MNL errors are $(-0.243, -0.025, 0.094)$, whereas for the MD errors, they are $(-0.172, 0.032, 0.387)$.

From these sensitivity experiments, we conclude that given product availability data and transaction data simulated from three commonly used, parametric, choice-based demand models, our MD algorithm can effectively infer a rank-based choice model from a completely uninformed prior about the underlying model. As expected, the variance of the error distribution is moderately larger under MD estimation. An implication of this is that MD requires a larger training data set to achieve more accurate estimates for $\mathbb{P}(j_t \mid S_t)$, as previously suggested by Table 2. Another important observation is that typically, the quality of our MD estimates compared with MLEs based on a perfect specification tends to improve as

Figure 6 (Color online) Relative Error of Probabilities Estimated by Maximum Likelihood Based on LC-MNL (Left) and the MD Algorithm (Right) with Respect to 50 Perturbed, Ground-Truth LC-MNL Models



the complexity of the demand model increases from MNL to CNL and LC-MNL.

5.2. Experiment Based on Real Hotel Data

We next present results applying the MD algorithm to a publicly available hotel data set (see Bodea et al. 2009). The booking records correspond to transient customers (predominately business travelers) with check-in dates between March 12, 2007, and April 15, 2007, in one of five continental U.S. hotels. For every hotel, a minimum booking horizon of four weeks for each check-in date was considered. Rate and room type availability information present at the time of booking was recorded for reservations made via the hotel or customer relationship officers, the hotels' websites, and off-line travel agencies. The data was preprocessed to comply with the model assumptions (e.g., there is at least one transaction per product). We define a *product* as a room type (e.g., king nonsmoking, queen smoking, etc), and a period as a (booking date, check-in date) pair. The original data set corresponds only to booking records, and therefore we set the arrival rate at $\lambda = 1$. Table 3 summarizes further details and the estimation results of the rank-based model relative to the independent demand and MNL models.

Given the limited amount of data for two of the hotels, our goodness-of-fit measures here are all in-sample. We report the AIC_c and the RMSE between the predicted and the observed bookings aggregated over the selling horizon. For the independent demand and the rank-based models, we also report the log-likelihood values (since they both correspond to the same log-likelihood function given in (3)).

In these examples, the single-class MNL model consistently dominates with respect to RMSE, with a remarkable performance. But the numbers also show an acceptable performance for the rank-based choice model. For instance, we note that for Hotel 1, the rank-based model dominates with respect to AIC_c, which indicates that this demand model provides a better compromise between goodness of fit and complexity than both MNL and the independent demand models. Note that this hotel is the one with more

Table 3 Estimation Results for the Hotel Example

Feature	Hotel 1	Hotel 2	Hotel 3	Hotel 4	Hotel 5
Number of products after preprocessing	10	12	8	9	8
Number of periods	1,315	211	1,147	288	245
Availability over selling horizon (%)	64	74	88	77	85
MNL					
AIC_c	4,131	606	2,359	735	722
RMSE	1.67	0.20	0.12	0.08	0.07
Independent demand					
Log-value	−2,621	−378	−1,569	−416	−406
AIC_c	5,262	782	3,155	851	822
RMSE	100.53	14.46	77.52	15.01	13.85
Rank based					
Number of final types	27	19	14	15	14
Log-value	−1,930	−289	−1,268	−358	−347
AIC_c	3,915	619	2,565	784	724
RMSE	7.66	4.72	8.50	4.97	3.52

data points and lower availability, and may be subject to more complex substitution behavior on the customer side, which is consistent with our observations in §5.1.2. The rank-based choice model also offers a very competitive performance in Hotels 2 and 5. In addition, the table shows the level of improvement that the MD procedure can bring to the independent demand model. By adding 17 types in the first hotel, and only between 6 and 7 types in the other hotels, we observe a quite significant drop in the RMSE (between 67% and 93%) and in the AIC$_c$ (between 8% and 26%), with a significant increase in the log-likelihood values (between 14% and 26%).

6. Extension: Market Discovery Under Censored Demand

As mentioned in §3, in settings such as bricks-and-mortar retailing, the modeler may not have access to shopping data and therefore cannot distinguish a period with no arrival from a period with an arrival that did not purchase (e.g., due to a stockout). This creates a new source of incompleteness in the data in addition to the nonobservability of the arriving customer type. Our MD algorithm can be adapted to account for this censored demand setting, albeit in a heuristic rather than in an exact way given the structure of the modified log-likelihood function.

Let $\mathcal{P}$ and $\bar{\mathcal{P}}$ denote, respectively, the set of periods with purchases and no purchases. In terms of the original log-likelihood formulation (e.g., see (2)), $\bar{\mathcal{P}} = \bar{\mathcal{P}}_\lambda \cup \bar{\mathcal{P}}_{\bar{\lambda}}$.

The incomplete data log-likelihood function for our problem is then given by

$$\begin{aligned}\mathcal{L}_I(\mathbf{x},\lambda) &= \sum_{t\in\mathcal{P}}\left(\log\lambda + \log\left(\sum_{i\in\mathcal{M}_t(j_t,S_t)} x_i\right)\right)\\ &\quad+\sum_{\substack{t\in\bar{\mathcal{P}},\\ \mathcal{M}_t(0,S_t)\neq\varnothing}}\log\left(\lambda\sum_{i\in\mathcal{M}_t(0,S_t)}x_i + (1-\lambda)\right)\\ &\quad+\sum_{\substack{t\in\bar{\mathcal{P}},\\ \mathcal{M}_t(0,S_t)=\varnothing}}\log(1-\lambda).\end{aligned}$$

The first term accounts for the likelihood of the observed transactions in the periods with purchases. The second term accounts for the no-purchase periods, where an arriving customer preferred not to buy (i.e., customer types in $\mathcal{M}_t(0, S_t)$) or no customer arrived at all. The third term accounts for the periods where none of the customer types would have picked a product from the available assortment, and hence $|\mathcal{M}_t(0, S_t)| = 0$. This last case indicates with certainty that no arrival occurred.

The MLE estimation problem (which extends (3)) can be formulated as follows:

$$\begin{aligned}\max_{\mathbf{x}\geq 0,\,0\leq\lambda\leq 1}\quad & \mathcal{L}_I(\mathbf{x},\lambda)\\ \text{s.t.}\quad & \textstyle\sum_{i=1}^K x_i = 1.\end{aligned}\tag{14}$$

This constrained, nonlinear optimization problem is hard to solve even for a fixed, given set of customer types. In fact, $\mathcal{L}_I(\mathbf{x}, \lambda)$ is no longer separable in $\mathbf{x}$ and λ, and it is not even quasi-concave in general, so we cannot preclude the possibility that there may exist local optima (see Proposition A1 in the online appendix).

Our proposal to compute MLE estimates extends the ideas developed for the uncensored demand case in a heuristic sense: we start from a limited set of types $\boldsymbol{\sigma} = \{\sigma^{(1)}, \ldots, \sigma^{(N)}\}$ and solve a restricted version of (14). This can be executed using a standard nonlinear optimization package or via an extension of the aforementioned EM procedure (see van Ryzin and Vulcano 2013) to obtain estimates $(\bar{\mathbf{x}}, \bar{\lambda})$ for the restricted problem. Then, we attempt to increase the log-likelihood value by adding a new type $\boldsymbol{\sigma}^{(N+1)}$ and repeat the procedure while we obtain a statistically significant increase of the (unrestricted) log-likelihood function.

Similar to the uncensored demand case, we define the matrix $A \in \mathcal{R}^{T\times K}$ with elements $a_{ti} = 1$ if $i \in \mathcal{M}_t(j_t, S_t)$ (here, j_t could also be zero) and $a_{ti} = 0$ otherwise, and we formulate the optimization problem:

$$\begin{aligned}\max_{\mathbf{x},\,\mathbf{y}\geq 0,\,\lambda\geq 0}\quad & \mathcal{L}_I(\mathbf{y},\lambda)\\ \text{s.t.}\quad & \sum_{i=1}^K x_i - 1 = 0,\\ & A\mathbf{x}-\mathbf{y} = 0,\\ & 1-\lambda\geq 0.\end{aligned}\tag{15}$$

We define the associated Lagrangian dual function:

$$\begin{aligned}\Theta(\beta,\boldsymbol{\mu},\gamma) = \sup_{\mathbf{x},\,\mathbf{y}\geq 0,\,\lambda\geq 0}\Bigg\{&\mathcal{L}_I(\mathbf{y},\lambda) + \beta\left(\sum_{i=1}^K x_i - 1\right)\\ &+\boldsymbol{\mu}^\mathsf{T}(A\mathbf{x}-\mathbf{y}) + \gamma(1-\lambda)\Bigg\}\end{aligned}$$

and consider the Lagrangian dual problem:

$$\min_{\beta,\,\boldsymbol{\mu},\,\gamma}\Theta(\beta,\boldsymbol{\mu},\gamma),\ \text{with}\ \beta\in\mathcal{R},\,\boldsymbol{\mu}\in\mathcal{R}^T,\,\gamma\in\mathcal{R}_+.$$

The analysis proceeds along the lines described in §4, with the caveat that now the strong duality theorem does not hold, and the gradient of the Lagrangian dual function only provides a heuristic improving direction to get a better primal solution.

Consider a restricted set of types $\boldsymbol{\sigma}$ and a stationary point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})$ of the restricted version of problem (15). Define

$$G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda) = \phi^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda, \bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma}),$$

$$\text{where } \phi^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda, \beta, \boldsymbol{\mu}, \gamma) = \mathcal{L}_I^{\boldsymbol{\sigma}}(\mathbf{y}, \lambda) + \beta\left(\sum_{i=1}^{N} x_i - 1\right) + \boldsymbol{\mu}^\mathsf{T}(A\mathbf{x} - \mathbf{y}) + \gamma(1-\lambda).$$

Assume that the KKT conditions hold at the point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda}, \bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma})$. Such a point should satisfy the following conditions for $\bar{\lambda}$ and $\bar{y}_t = \sum_{i \in \mathcal{M}_t(j_t, S_t)} \bar{x}_i$:

- For $t \in \mathcal{P}$,

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda)}{\partial y_t}\bigg|_{(\mathbf{x}, \mathbf{y}, \lambda) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})} = \frac{1}{\bar{y}_t} - \bar{\mu}_t = 0,$$

$$\text{and then } \bar{\mu}_t = \frac{1}{\bar{y}_t}.$$

- For $t \in \bar{\mathcal{P}}$, $\mathcal{M}_t(0, S_t) \neq \varnothing$,

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda)}{\partial y_t}\bigg|_{(\mathbf{x}, \mathbf{y}, \lambda) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})} = \frac{\bar{\lambda}}{\bar{\lambda}\bar{y}_t + 1 - \bar{\lambda}} - \bar{\mu}_t = 0,$$

$$\text{and then } \bar{\mu}_t = \frac{\bar{\lambda}}{\bar{\lambda}\bar{y}_t + 1 - \bar{\lambda}}.$$

- For $t \in \bar{\mathcal{P}}$, $\mathcal{M}_t(0, S_t) = \varnothing$,

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda)}{\partial y_t}\bigg|_{(\mathbf{x}, \mathbf{y}, \lambda) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})} = -\bar{\mu}_t = 0,$$

$$\text{and then } \bar{\mu}_t = 0.$$

- The value $\bar{\beta}$ can be computed from the derivative with respect to any x_i:

$$\frac{\partial G^{\boldsymbol{\sigma}}(\mathbf{x}, \mathbf{y}, \lambda)}{\partial x_i}\bigg|_{(\mathbf{x}, \mathbf{y}, \lambda) = (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})} = \bar{\beta} + \sum_{t:\ i \in \mathcal{M}_t(j_t, S_t)} \bar{\mu}_t = 0 \quad \text{for all } i.$$

An analysis similar to the uncensored demand case gives

$$\bar{\beta} = -\left(|\mathcal{P}| + \sum_{\substack{t \in \bar{\mathcal{P}} \\ \mathcal{M}_t(0, S_t) \neq \varnothing}} \frac{\bar{\lambda}\bar{y}_t}{\bar{\lambda}\bar{y}_t + 1 - \bar{\lambda}}\right).$$

Consider again the full set of types and the Lagrangian dual function

$$\Theta(\bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma}) = \sup_{\mathbf{x}, \mathbf{y} \geq 0,\ \lambda \geq 0} \{\phi(\mathbf{x}, \mathbf{y}, \lambda, \bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma})\}.$$

Define $G(\mathbf{x}, \mathbf{y}, \lambda) = \phi(\mathbf{x}, \mathbf{y}, \lambda, \bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma})$. If for all i, $N+1 \leq i \leq K$, $\partial G(\mathbf{x}, \mathbf{y}, \lambda)/\partial x_i \leq 0$, then $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda}, \bar{\beta}, \bar{\boldsymbol{\mu}}, \bar{\gamma})$ is a saddle point of the Lagrangian, and $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\lambda})$ is a stationary point of primal problem (15) (and hence of original problem (14)). Otherwise, there exists k, $N+1 \leq k \leq K$, such that

$$\frac{\partial G(\mathbf{x}, \mathbf{y}, \lambda)}{\partial x_k} = \sum_{\substack{t:\ \sigma^{(k)}(j_t) < \sigma^{(k)}(i) \\ \forall i \in S_t,\ i \neq j_t}} \bar{\mu}_t > -\bar{\beta}.$$

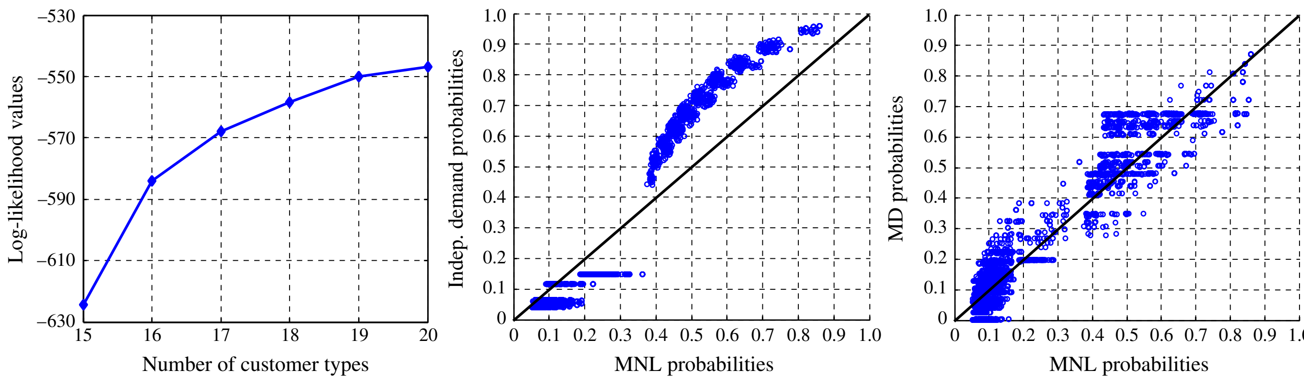
Customer type k defines a direction of improvement for the Lagrangian, which *may* lead to an improved primal solution. Our implementation still controls for a net statistically significant improvement of the log-likelihood value and, under this heuristic approach, also to prevent a potential worsening of the primal solution.

We illustrate the MD algorithm under a censored demand model by revisiting the Amazon MNL model presented in §5. We assume an underlying arrival rate $\lambda = 0.9$ and generate 500 periods of transactions for the training data set and 1,000 periods for the out-of-sample data set. Under this censored demand case, the market share is nonobservable. Indeed, a major advantage of the MD algorithm is that it does not require market share information, which is a common input parameter for other models (including MNL, CNL, and LC-MNL). Figure 7 exhibits the results on 4,982 out-of-sample data points. The general pattern is similar to what was observed in Figure 1. In this case, the MD algorithm adds five types to the basic set of singletons (see Figure 7, left panel). The scatterplot in the center panel depicts the optimal independent demand probabilities versus the true MNL probabilities. The RMSE between the estimated probabilities and the true probabilities is 0.09, and the AIC$_c$ is 1,278.7. The independent demand-based arrival rate slightly overestimates the true one at 0.97. The scatterplot in the right panel depicts the predicted rank-based probabilities versus the true MNL probabilities. The RMSE decreases to 0.06, and the AIC$_c$ decreases to 1,134.2. Even though the arrival rate is underestimated at 0.73 in this case, the predicted probabilities are better aligned around the 45° line, indicating again the improvement of the estimated probabilities from our MD algorithm.

7. Conclusions

In this paper, we propose a procedure to inferring customer preferences for a set of substitutable products using only sales transactions and product availability data. The demand model is defined by a Bernoulli process of arrivals over time, where each arrival chooses among the set of available products according to preference rankings of alternatives (including the no-purchase alternative). Our approach jointly estimates the arrival rate and the pmf of the rank-based choice model under a maximum likelihood criterion. To overcome the problem of the factorial number of rankings of products, we develop an iterative market discovery algorithm that consists of

Figure 7 (Color online) Performance of the MD Algorithm When Facing an Underlying MNL Model with Censored Arrivals



Notes. Left, evolution of the log-likelihood function within MD. Center, predicted probabilities under the independent demand assumption versus ground-truth probabilities. Right, predicted probabilities under MD versus ground-truth probabilities.

a sequence of *estimation* and *discovery* stages. Starting from a parsimonious set of rankings (i.e., customer types), we compute maximum likelihood estimates for these types and then use dual information from the estimation phase to formulate an MIP that generates a new customer type that increases the likelihood value.

Given its ability to automate the specification of the set of customer types in cases with complex choice sets that vary over time, the realistic and limited input data needed, and the quality of the results, we believe that our proposal has significant practical potential.

Our numerical experiments also provide some preliminary evidence that as the underlying choice model becomes more complex, rank-based models produced by our procedure provide good predictive results relative to several common parametric choice models. Still, it would be worth exploring this issue in a more comprehensive study to better understand the relative strengths and limitations of parametric and nonparametric approaches to modeling choice behavior and when one approach might be preferred over the other.

The following two extensions further enhance its real-world viability. In van Ryzin and Vulcano (2013), we develop a simple and fast EM algorithm for the estimation phase that is more efficient than direct maximization of the incomplete data log-likelihood function. The paper by Méndez-Díaz et al. (2014) leverages the column generation procedure by proposing a branch-and-cut that significantly reduces its computational burden.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.2040.

Acknowledgments

The authors thank Richard Ratliff (Sabre Holdings) and Martín González-Rozada and Martín Solá (Universidad Torcuato di Tella, Argentina) for their helpful suggestions on earlier drafts of this work. The authors are also grateful to the associate editor and three anonymous referees for their constructive feedback.

References

Andersson S-E (1998) Passenger choice analysis for seat capacity control: A pilot project in Scandinavian Airlines. *Internat. Trans. Oper. Res.* 5(6):471–486.

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* 17(4):406–423.

Barberá S, Pattanaik PK (1986) Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica* 54(3):707–715.

Bazaraa M, Sherali H, Shetty C (2006) *Nonlinear Programming: Theory and Algorithms*, 3rd ed. (John Wiley & Sons, Hoboken, NJ).

Ben-Akiva M, Lerman S (1994) *Discrete Choice Analysis: Theory and Applications to Travel Demand*, 6th ed. (MIT Press, Cambridge, MA).

Block H, Marschak J (1960) Random orderings and stochastic theories of responses. Olkin I, Ghurye S, Hoeffding W, Madow W, Man H, eds. *Contributions to Probability and Statistics* (Stanford University Press, Redwood City, CA), 97–132.

Bodea TD, Ferguson ME, Garrow LA (2009) Choice-based revenue management: Data from a major hotel chain. *Manufacturing Service Oper. Management* 11(2):356–361.

Borle S, Boatwright P, Kadane JB, Nunes JC, Galit S (2005) The effect of product assortment changes on customer retention. *Marketing Sci.* 24(4):616–622.

Bruno HA, Vilcassim NJ (2008) Structural demand estimation with varying product availability. *Marketing Sci.* 27(6):1126–1131.

Chaneton JM, Vulcano G (2011) Computing bid-prices for revenue management under customer choice behavior. *Manufacturing Service Oper. Management* 13(4):452–470.

Chen L, Homem-de-Mello T (2010) Mathematical programming models for revenue management under customer choice. *Eur. J. Oper. Res.* 203(2):294–305.

Chintagunta PK, Dubé J-P (2005) Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data. *J. Marketing Res.* 42(3):368–379.

Cooper WL, Homem-de-Mello T, Kleywegt AJ (2006) Models of the spiral-down effect in revenue management. *Oper. Res.* 54(5):968–987.

Falmagne JC (1978) A representation theorem for finite random scale systems. *J. Math. Psych.* 18(1):52–72.

Farias V, Jagabathula S, Shah D (2012) Assortment optimization under general choice. Working paper, MIT Sloan School of Management, Cambridge, MA.

Farias VF, Jagabathula S, Shah D (2013) A nonparametric approach to modeling choice with limited data. *Management Sci.* 59(2): 305–322.

Garey MR, Johnson DS (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman & Company, San Francisco).

Greene W (2003) *Econometric Analysis*, 5th ed. (Prentice Hall, Englewood Cliffs, NJ).

Gruen T, Corsten D, Bharadwaj S (2002) Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses. Report for the Grocery Manufacturers of America, Washington, DC.

Haensel A, Koole G (2011) Estimating unconstrained demand rate functions using customer choice sets. *J. Revenue Pricing Management* 10(5):438–454.

Honhon D, Jonnalagedda S, Pan XA (2012) Optimal algorithms for assortment selection under ranking-based consumer choice models. *Manufacturing Service Oper. Management* 14(2): 279–289.

Ja S, Rao BV, Chandler S (2001) Passenger recapture estimation in airline revenue management. *AGIFORS 41st Annual Sympos., Sydney, NSW, Australia*.

Kalyanam K, Borle S, Boatwright P (2007) Deconstructing each item's category contribution. *Marketing Sci.* 26(3):327–341.

Kök AG, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Oper. Res.* 55(6):1001–1021.

Kunnumkal S (2014) Randomization approaches for network revenue management with customer choice behavior. *Production Oper. Management* 23(9):1617–1633.

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Oper. Res.* 49(3):334–351.

Manksi CF (1977) The structure of random utility models. *Theory Decisions* 8(3):229–254.

Manksi CF (2007) Partial identification of counterfactual choice probabilities. *Internat. Econom. Rev.* 48(4):1393–1410.

Martí R, Reinelt G (2011) *The Linear Ordering Problem: Exact and Heuristic Methods in Combinatorial Optimization*, 1st ed. (Springer, Berlin).

Méndez-Díaz I, Vulcano G, Zabala P (2014) A branch-and-cut algorithm for a generalized linear ordering problem. Working paper, Leonard N. Stern School of Business, New York University, New York.

Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Sci.* 56(7):1180–1197.

Newman JP, Ferguson ME, Garrow LA, Jacobs TL (2014) Estimation of choice-based models using sales data from a single firm. *Manufacturing Service Oper. Management* 16(2):184–197.

Queenan CC, Ferguson M, Higbie J, Kapoor R (2007) A comparison of unconstraining methods to improve revenue management systems. *Production Oper. Management* 16(6):729–746.

Ratliff RM, Rao BV, Narayan CP, Yellepeddi K (2008) A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *J. Revenue Pricing Management* 7(2):153–171.

Rusmevichientong P, Van Roy B, Glynn PW (2006) A nonparametric approach to multiproduct pricing. *Oper. Res.* 54(1):82–98.

Sher I, Fox J, Kim K, Bajari P (2011) Partial identification of heterogeneity in preference orderings over discrete choices. NBER Working Paper 17346, National Bureau of Economic Research, Cambridge, MA.

Swait J, Erdem T (2002) The effects of temporal consistency of sales promotions and availability on consumer choice behavior. *J. Marketing Res.* 39(3):304–320.

Talluri KT (2009) A finite-population revenue management model and a risk-ratio procedure for the joint estimation of population size parameters. Working paper, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain.

Talluri K, van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* 50(1):15–33.

Train K (2003) *Discrete Choice Methods with Simulation* (Cambridge University Press, New York).

van Ryzin G, Vulcano G (2008) Computing virtual nesting controls for network revenue management under customer choice behavior. *Manufacturing Service Oper. Management* 10(3): 448–467.

van Ryzin G, Vulcano G (2013) An expectation-maximization method to estimate a rank-based choice model of demand. Working paper, Decision, Risk and Operations Division, Columbia Business School, New York.

Vulcano G, van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.* 60(2):313–334.

Zhang D, Cooper WL (2006) Revenue management for parallel flights with customer-choice behavior. *Oper. Res.* 53(3):415–431.

Zuckerman D (2007) Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory Comput.* 3:103–128.