

HIER

Harvard Institute of Economic Research

Discussion Paper Number 2174

Demand Estimation Under Incomplete
Product Availability

by

Christopher T. Conlon
and
Julie Holland Mortimer

May 2009

Harvard University
Cambridge, Massachusetts

This paper can be downloaded without charge from:
<http://www.economics.harvard.edu/journals/hier2009>

Demand Estimation Under Incomplete Product Availability*

Christopher T. Conlon[†]
Julie Holland Mortimer[‡]

First version: October 24, 2006
March 16, 2009

Abstract

Incomplete product availability arising from stock-out events and capacity constraints is a common and important feature of many markets. Periods of unavailability censor the observed sales for the affected product, and potentially increase observed sales of available substitutes. As a result, failing to adjust for incomplete product availability can lead to biased demand estimates. Common applications of these demand estimates, such as computing welfare effects from mergers or new products, are therefore unreliable in such settings. These issues are likely to arise in many industries, from retail to sporting events to airlines. In this paper, we study a new and unique dataset from one of the first wireless inventory management systems in the country, which was installed on a set of 54 vending machines in order to track product availability at high frequency (roughly every four hours). These data allow us to account for product availability when estimating demand, and introduces a valuable source of variation for identifying substitution patterns. We also develop a simple procedure that allows for changes in product availability even when we only observe inventory (and thus availability) periodically. We find significant differences in the parameter estimates of demand, and as a result, the corrected model predicts significantly larger impacts of stock-outs on profitability.

*We thank Dan Akerberg, Susan Athey, Steve Berry, Uli Doraszelski, J.P. Dube, Phil Haile, Wes Hartmann, Ken Hendricks, Guido Imbens, Phillip Leslie, Richard Mortimer, Ariel Pakes and Frank Wolak for helpful discussions and comments. We also thank seminar participants at the Stanford Institute for Theoretical Economics, the National Bureau of Economic Research (Summer Institute), the Econometric Society (Winter Meetings), Columbia University, Harvard University, London School of Economics, the Stern School of Business at New York University, Stanford University, University of Arizona, and Washington University for helpful comments. Financial support for this research was generously provided through NSF grant SES-0617896.

[†]Yale University, Department of Economics, 37 Hillhouse Ave., New Haven, CT, 06511. email: christopher.conlon@yale.edu

[‡]Harvard University and NBER, Department of Economics, 1805 Cambridge St., Cambridge, MA, 02138. email: mortimer@fas.harvard.edu

1 Introduction

Incomplete product availability is a common and important feature of markets where products are perishable, seasonal, or have storage costs. For example, retail markets, sporting and concert events, and airlines face important capacity constraints that often lead to stock outs. Not surprisingly, firms in such industries identify inventory management as a critical strategic decision, and consumers cite product availability as a major concern.¹ In these settings, the failure to account for product availability not only ignores a useful source of variation for identifying demand parameters, but can also lead to biased estimates of demand. The first source of bias is the *censoring* of demand estimates. If a product sells out, the actual demand for a product (at given prices) may be greater than the observed sales, leading to a negative bias in demand estimates. At the same time, during periods of reduced availability of other products, sales of available products may increase. This *forced substitution* overstates demand for these goods conditional on the full choice set being available. As a result, failing to account for product availability leads to biased estimates of demand substitution patterns, typically making products look more substitutable than they really are. This bias can potentially undermine the reliability of many important applications of demand estimates for markets with incomplete product availability, such as simulating the welfare implications of mergers or new product introductions, or applying antitrust policy. Identifying unbiased demand estimates in these markets is also a critical step in evaluating optimal capacity choices of firms.

In this paper, we provide evidence that failing to appropriately account for periods of product unavailability can result in a substantial bias in demand estimates, and we develop a method for correcting this bias. To accomplish this, we collected a new and extensive dataset with detailed inventory and sales information. The dataset covers one of the first technological investments for wirelessly managing inventory: a wireless network installed on a set of 54 vending machines, providing updates on elapsed sales and inventory status every four hours. The data from the vending network provide extremely granular information on sales and inventory levels over a period of a year. Using this dataset, we develop and implement estimation methods to provide corrected estimates even when some choice sets are latent, and analyze the impact of stockouts for firm profitability in the short run. We find evidence of important biases on demand parameters and predicted sales in models that do not account correctly for stock-out events in this market. For example, under some specifications of the uncorrected model, we estimate demand parameters that are not consistent with utility maximization. In terms of the short-run impacts of stock-outs on profitability, the corrected model estimates that the negative profit impacts of stock-outs are 8 - 12 percent larger than the uncorrected model predicts.

Although not estimated here, the model we develop is also necessary for any examination of supply-side decisions over the long run. For example, estimation of optimal capacity

¹Relatedly, firms throughout the economy are currently making large investments in technologies for tracking inventory and capacity information. For example, Walmart now requires many suppliers to use Radio Frequency Identification (RFID) technology, and many other firms have recently adopted related technology, such as wireless communication and networked data centers.

choices, restocking decisions or inventory policies in markets where stock-outs matter relies on a static demand model that accounts correctly for stock-out events as our model does.² Demand estimates that correctly account for product availability are also important for understanding the macroeconomic implications of inventories. Indeed, firms’ abilities to manage inventories have been proposed as an agent for dampening recessions, a factor affecting vertical relationships, and a strategic variable affecting price competition.³

The wireless vending network provides actual stock-out events, which randomly change the set of products available at some locations for a period of time. This variation provides an attractive source of identification for estimating demand models, because although the probability of a stock-out event can be targeted with different inventory choices, the occurrence of any particular event is not chosen by the firm. Thus, stock outs generate exogenous “short-run” variation in choice sets, in addition to the long-run variation that is more typically the source of identification for structural models of demand. Many other data sources have the potential to generate similar sources of variation in availability, which makes the exercise here potentially quite broadly applicable. Any market for which the data contain repeated draws of demand and we observe capacity generates similar variation in choice sets. This variation comes from periods in which demand exceeds capacity constraints, and we investigate and example of this phenomenon with respect to vending data.

When discussing inventory systems we use the standard language established by Hadley and Whitman (1963). The first of two types of inventory systems is called a ‘perpetual’ data system. In this system, product availability is known and recorded when each purchase is made. Thus for every purchase, the retailer knows exactly how many units of each product are available.⁴ The other type of inventory system is known as a ‘periodic’ inventory system. In this system, inventory is measured only at the beginning of each period. After the initial measurement, sales take place, but inventory is not measured again until the next period. Periodic inventory systems are problematic in analyses of stock-outs because inventory (and thus the consumer’s choice set) is not recorded with each transaction. While perpetual inventory systems are becoming more common in retailing environments, most retailers still do not have access to such systems. Sampling inventory more frequently helps to mitigate limitations of the periodic inventory system. However, an additional goal of this paper is to provide consistent estimates of demand not only for perpetual inventory systems but for periodic ones as well.

In fact, despite the extremely detailed information from this dataset, we observe stock-

²Such supply-side problems require an additional focus on firm costs and dynamic inventory decisions, and we analyze such a model in a companion paper with Uli Doraszelski (in progress).

³McCarthy and Zakrajsek (forthcoming) reviews the literature on the effect of inventory management technology on business cycles, and provides empirical evidence on the theory. Narayanan and Raman (2004) examines the assignment of stocking rights in vertical settings theoretically, and Mortimer (2008) provides empirical evidence on the effect of inventory monitoring technology for vertical contracting in the video rental industry. Balachander and Farquhar (1994), Carlton (1978), Dana (2001), and Deneckere and Peck (1995), among others, address the impact of product availability on price or service competition.

⁴Note that if sales are recorded in the order they happen, this would be sufficient to construct an almost perpetual inventory system (assuming consumers do not hold goods for long before purchasing an item). This system is also known as ‘real-time’ inventory.

out events only periodically (every four hours). Some stock-out events occur in the middle of an observed four-hour time period, meaning that for these observations, the choice set of an individual consumer is latent. Discarding data from these periods would select on sales levels and lead to biased estimates. Thus, we develop a method for incorporating these observations that uses the well-known EM algorithm from the statistics literature (Dempster, Laird, and Rubin 1977) to estimate the allocation of sales across the unobserved choice set regimes.

As technologies like the one we study continue to become more prevalent, firms and researchers can expect to gain access to better data (i.e., more detailed information on sales and inventory/capacities) with which to analyze markets. Note that new technologies are not always necessary to observe such information; capacity data provide the same functionality for many markets, such as airlines, performance or sporting events, hospitals, etc. When such data are available, researchers gain valuable information on short-run choice set variation. Our results in this paper indicate that accounting for that choice set variation can substantially reduce potential biases in standard estimates for some markets, and that researchers should take on the responsibility to adjust for the effects of product availability in demand estimation when possible.

Relationship to Literature

The demand estimation literature in IO has primarily focused on cases in which all products are assumed to be available to all consumers ((Berry, Levinsohn, and Pakes 1995, Nevo 2001, Berry, Levinsohn, and Pakes 2004)). Dynamic models of demand ((Nevo and Hendel 2007)) have modeled consumer inventories or stockpiling, but do not incorporate data on availability at the time of purchase. Data on capacities/inventories have been used primarily in this literature in order to analyze supply-side behavior—particularly with respect to promotions ((Aguirregabiria 1999)).⁵

The marketing literature examines problems that arise when data are aggregated over the choices of consumers facing different choice sets, which is another way of viewing the problem of product availability. This literature points out some of the same biases we examine here. For example, Gupta, Chintagunta, Kaul, and Wittink (1995) review biases in city-level demand that arise when different stores carry different products and have different promotional conditions. In another approach, Bruno and Vilcassim (2008) have data that are aggregated across choice sets and assume additional structure on the marginal distributions of availability across retail outlets within a market to probability weight to different choice sets.

Stock-outs are frequently analyzed in the context of optimal inventory policies in operations research. In fact, an empirical analysis of stock-out based substitution has been addressed using vending data before by Anupindi, Dada, and Gupta (1998) (henceforth ADG). ADG use an eight-product soft-drink machine and observe the inventory at the beginning of each day. The authors assume that products are sold at a constant Poisson distributed rate (cans per hour). The sales rates of the products are treated as independent from one

⁵While we explicitly account for retailer inventories in our model, dynamic supply-side behavior does not arise because the retailer does not have the ability to dynamically alter the price or product mix.

another, and eight Poisson parameters are estimated. When a stock-out occurs, a new set of parameters is estimated with the restriction that the new set of parameters are at least as great as the original parameters. This means that each choice set requires its own set of parameters (and observed sales). If a Poisson rate was not fitted for a particular choice set, then only bounds can be inferred from the model. However, because of the lack of a utility-based framework for demand, the ADG method cannot be used to make out-of-sample predictions about alternative policies or their welfare impacts. More recently, Musalem, Olivares, Bradlow, Terwiesch, and Corsten (2008) study the problem of stock-outs by imputing the entire sequence of sales in a Markov chain Monte Carlo approach. The MLE approach that we use is more efficient and faster to implement, particularly for markets in which the outside good has a relatively large share and the total number of products stocked-out during period of latent availability is relatively small.

The economics literature has also studied Bayesian models of discrete choice consumer demand, for example, Athey and Imbens (2007). While our paper uses a common Bayesian technique to address missing data, it is not a fully Bayesian model. Relatedly, Fox (2007) examines semiparametric estimation of multinomial discrete choice models using a subset of choices, and proposes a maximum score estimator. His focus includes settings with very large choice sets, for which estimation is computationally burdensome, as well as choice subsets that arise due to data limitations or consideration sets. However, this methodology does not allow for computation of market shares, which limits the ability to analyze the welfare effects of stockouts.

The paper proceeds as follows. Section 2 provides the model of demand for finely observed data, and section 3 adjusts for changes in product availability in the data under both perpetual and periodic inventory systems. In section 4 we provide estimation details and discuss identification of the model. Section 5 describes the data from the wireless vending route and provides correlations and regression results from the data. Section 6 reports results from estimating the model using the vending data, section 7 provides counterfactual experiments on the effect of stockouts on firm profitability, and section 8 concludes.

2 Model

In this section, we develop a model of consumer sales. We start with a multinomial distribution for any general set of consumer choice probabilities (denoted p_{ij} for consumer i and product j) rather than starting with a specification of p_{ij} (such as logit) and using this as an input into a multinomial distribution. We use this focus in order to present the impact of stockouts on the estimator in a clear way. Once we describe the model and the method for incorporating information on stockout events, we adopt the usual logit/random-coefficient logit specification of consumer choice probabilities in the estimation section.

We denote the purchase of consumer i as y_i , and let x_i denote the relevant observables. Typically we think of y_i as a categorical variable taking one of several discrete values $j = (0, 1, \dots, J)$. In an abuse of notation we also let $y_{ij} = 1$ for the chosen product and $y_{ik} = 0$ for products that are not chosen. Thus y is both an index for which product was chosen as well

as an indicator for the chosen product. For simplicity, we denote the set of possible choices as J , the powerset of $\{0, 1, \dots, J\}$ as A , and $a_i \in A$ as the set of products available to consumer i , always including the outside good. Continuing the standard abuse of notation, we'll also consider a to be a $J \times 1$ vector that takes on value 1 in the j th position if the product is available and zero otherwise. We assume that each consumer chooses some product $j \in a_i$ or the outside good $j = 0$.

Without making any parametric assumptions we can write down a multinomial likelihood of seeing an individual choose product j as:

$$L(y_i|\theta) \propto \prod_{j \in a_i} p_{ij}^{y_i=j} = p_{ij}$$

because $y_i = 1$ for the observed choice and zero elsewhere. We can consider the joint likelihood of observing y_1, \dots, y_n , which is the product of the probabilities of the observed choices of $i = 1, \dots, n$ consumers.

$$L(y_1, \dots, y_n|\theta) \propto \prod_{i=1}^n \prod_{j \in J} p_{ij}^{y_{ij}} = \prod_{i=1}^n p_{ij}$$

$$l(y_1, \dots, y_n|\theta) \propto \sum_{i=1}^n \sum_{j \in J} y_{ij} \ln p_{ij} = \sum_{i=1}^n \ln p_{ij}$$

2.1 Sufficient Statistics

Typically, researchers specify choice probabilities as $p_{ij} = p_j(\Phi_i, \theta)$.⁶ Thus, an individual's probability of choosing product j depends on some observable Φ_i as well as the unknown parameters. The Φ_i could include information about the set of products available to consumer i (a_i), and information about the situation under which consumer i 's choice took place, such as time of day or location (x_i).

When we have data in which multiple individuals are described by the same Φ_i , then we can combine purchase decisions from these individuals without loss. That is, we can construct a new index t which sums over all possible values of Φ_i :

⁶This is the probability of choosing j conditional on Φ_i, θ .

$$\begin{aligned}
\sum_{i=1}^n \sum_{\forall j \in J} y_{ij} \ln p_{ij}(\Phi_i, \theta) &= \sum_{t=1}^T \sum_{\forall j \in J} \underbrace{\left(\sum_{i: \Phi_i = \Phi_t} y_{ij} \right)}_{y_{jt}} \ln p_j(\Phi_t, \theta) \\
&= \sum_{t=1}^T \sum_{\forall j \in J} y_{jt} \ln p_j(\Phi_t, \theta)
\end{aligned}$$

Thus y_{jt} is defined as the aggregate purchases of consumers facing the conditions described by Φ_t , and is a sufficient statistic for estimation. That is, y_{jt} contains all of the necessary information about y_{ij} that we need in order to compute the likelihood.

2.2 Aggregation

We often observe only aggregate data for some period t , denoted $y_{jt} = \sum_{i \in t} y_{it}$. If every consumer in period t faces the same Φ_i then we can consider such aggregate data without loss, as above. That is, one can write the likelihood function as a single multinomial with parameters $n = \sum_{j=0}^J y_{jt}$, and $p(\Phi_t, \theta) = [p_{1t}(\Phi_t, \theta), p_{2t}(\Phi_t, \theta), \dots, p_{Jt}(\Phi_t, \theta)]$ rather than a product of N multinomials with $n = 1$:

$$\begin{aligned}
f(\mathbf{y}_t | \theta, \Phi_t) &= \left(\frac{(\sum_{j=0}^J y_{jt})!}{y_{0t}! y_{1t}! y_{2t}! \dots y_{Jt}!} \right) p_{0t}(\Phi_t, \theta)^{y_{0t}} p_{1t}(\Phi_t, \theta)^{y_{1t}} \dots p_{Jt}(\Phi_t, \theta)^{y_{Jt}} \\
&= C(\mathbf{y}_t) p_{0t}(\Phi_t, \theta)^{y_{0t}} p_{1t}(\Phi_t, \theta)^{y_{1t}} \dots p_{Jt}(\Phi_t, \theta)^{y_{Jt}} \\
&\propto p_{0t}(\Phi_t, \theta)^{y_{0t}} p_{1t}(\Phi_t, \theta)^{y_{1t}} \dots p_{Jt}(\Phi_t, \theta)^{y_{Jt}} \tag{1}
\end{aligned}$$

An important simplification arises from the fact that the combinatorial term $C(\mathbf{y}_t)$ (which addresses the fact that the ordering is unobserved) depends only on the data, and does not vary with the parameter θ .

In order to consider aggregate data as if they were individual purchases requires that for all consumers $i \in t$, $\Phi_i = \Phi_t$ or $(a_i, x_i) = (a_t, x_t)$. For simplicity, let $\mathbf{y}_t = [y_{0t}, y_{1t}, y_{2t}, \dots, y_{Jt}]$. Then for each market, the data provide information on (\mathbf{y}_t, a_t, x_t) .

$$l(\mathbf{y}_t | \theta, a_t, x_t) = \sum_{j \in a_t} y_{jt} \ln p_j(x_t, a_t, \theta) = \sum_{j \in a_t} \sum_{i \in t} y_{it} \ln p_j(x_t, a_t, \theta)$$

The key is that when Φ_i is known and fixed at a level of aggregation, the sufficient statistics and hence the likelihood are invariant to the ordering of the observations, and we can treat aggregate data as if it were individual purchase data. To formalize this assumption:

Assumption 1. (*Conditional Independence of Choice Probabilities*) Up to an unknown parameter θ , Φ contains all information relevant for calculating consumer choice probabilities $p_{ij}(\Phi, \theta)$ and hence the likelihood.

It is important to understand the difference between the assumption of Conditional Independence and the more typical IID assumption. For example, in the model we consider (a logit form for $p_j(\cdot)$), choice probabilities depend on which products are available, a_i , but this is correlated with previous sales when stockouts are considered. This clearly violates the assumption of y_i being IID, since it is now distributed differently after the stockout in a way that clearly depends on the previous y_i realizations. However, this sequence is conditionally independent, because choice probabilities depend on previous sales only through $\Phi_i = (a_i, x_i)$. Hence if we know a_i and can condition on it, then the ordering of the y_i 's no longer affects the likelihood.

When using market-level data in general, we rule out cases where previous sales affect demand for products in ways other than through product availability. For example, we rule out psychological effects of other individuals' purchases on the purchase probability of subsequent consumers. We also rule out dynamic effects, such as the presence of pent-up demand following a stockout unless these events are captured in the x_i .

Unobserved heterogeneity, such as in random-coefficients models, does not violate the Conditional Independence assumption because we can compute the expected choice probabilities, $E[p_{ij}(\Phi, \theta)]$, over a distribution of tastes. However, if consumers are drawn from different taste distributions within a market, one can only make inferences about the overall mixture of these distributions.⁷ Thus, as in our model, these models also require that the population of consumer preferences cannot change within a level of aggregation, denoted as t . Without this assumption in our setting, one cannot tell apart stock-out events from atypical consumers. For example, if we observe data on sales between 4pm and 8pm, and the population of consumers changes at 5pm, we can't necessarily draw conclusions about the different preferences of the two consumer groups, but we can estimate the overall distribution of preferences in the population. Such latent types can create problems if we believe an x is an important determinant of choice probabilities, but it is partially or fully unobserved.⁸

⁷ Previous studies have relied on annual or quarterly data, (Berry, Levinsohn, and Pakes 1995, Nevo 2001) for which short-term heterogeneity in the population similarly gets "averaged out" in the overall distribution of consumer preferences.

⁸It is possible to handle these scenarios with some extensions to the method we describe. One method for accommodating latent types is to write down a mixture form for choice probabilities which integrates out the latent variable. We describe this approach in section A.1 of the appendix.

3 Adjusting for Choice Set Heterogeneity

3.1 Perpetual Inventory/Observed Choice Set Heterogeneity

We now consider the case where availability is observed for all sales (the case of perpetual inventory or sequential sales data with constant capacity). We relax the assumption that a_t (the set of available products) is constant across a time period, and allow for availability to change in the middle of a period t . In this case, the ‘period’ can be divided into two smaller periods of constant availability, denoted (a_s, a_t) and we recompute the sufficient statistics as follows:

$$\begin{aligned}
 l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \left(\sum_{\forall j \in a_s} \ln p_j(\theta, a_s, x_t) \sum_{\forall i: (a_{it}, x_{it}) = (a_s, x_t)} y_{ji} + \sum_{\forall j \in a_t} \ln p_j(\theta, a_t, x_t) \sum_{\forall i: (a_{it}, x_{it}) = (a_t, x_t)} y_{ji} \right) \\
 l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \left(\sum_{\forall (a, x)} \sum_{\forall j \in a} \ln p_j(\theta, a, x) \sum_{\forall i: (a_i, x_i) = (a, x)} y_{ji} \right) \\
 &= \sum_{\forall t} \left(\sum_{\forall (a, x)} \sum_{\forall j \in a} y_{j, (a, x)} \ln p_j(\theta, a, x) \right) \tag{2}
 \end{aligned}$$

3.2 Periodic Inventory/ Latent Choice Set Heterogeneity

In many market settings, firms only observe inventories or sales periodically. This presents additional challenges, because availability is known only at the beginning and the end of each period. As in the case of perpetual inventory, we could denote the set of available choices at the beginning of period t by a_s , and the set remaining at the end of t by a_t . One would like to assign the sales in period t to each (a_s, a_t) regime, but because inventory is observed only periodically, this is not possible. A standard approach for dealing with unobservable heterogeneity is to integrate out the heterogeneity and to work with the expectation instead. In other words, we cannot solve:

$$\hat{\theta} = \arg \max_{\theta} l(\mathbf{y}|\mathbf{a}, \mathbf{x}, \theta) = \sum_{i=1}^n \sum_{j \in a_i} y_{ij} \ln p_j(a_i, x_i, \theta)$$

because the vector of sales across all possible availability regimes is not fully observable for some periods. We partition the data into fully observable and partially observable periods, which we denote T_{mis} and T_{obs} for periods of unknown and known availability, respectively. For periods in T_{obs} , we know the availability regime that prevailed for all sales. Thus, we only need an expectation of sales for periods in T_{mis} . We re-write the expected log-likelihood function to account for both sets of periods as follows.

$$l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) \propto \sum_{\forall i} y_{ij} \ln p_j(\theta, a_i, x_i) = \sum_{\forall i \in T_{obs}} y_{ij} \ln p_j(\theta, a_i, x_i) + \sum_{\forall i \in T_{mis}} y_{ij} \ln p_j(\theta, a_i, x_i)$$

from equation 2, when $i \in T_{obs}$, (y_{ij}, a_i, x_i) are all known, and the sufficient statistic for estimation is

$$y_{j,(a,x)}^{obs} = \sum_{i \in T_{obs}: (a_i, x_i) = (a, x)} y_{ij} \quad (3)$$

For $i \in T_{mis}$, we don't know the value of y_{ij} but we know the possible (a, x) regimes. Thus, for any possible (a, x) pair, we replace $y_{j,(a,x)}^{mis}$ with its expectation, and re-write the log-likelihood as:

$$\sum_{\forall (a,x)} \left(\sum_{\forall t \in T_{mis}} E[y_{j,(a,x)}^{mis}] \ln p_j(\theta, a, x) + \sum_{\forall t \in T_{obs}} y_{j,(a,x)}^{obs} \ln p_j(\theta, a, x) \right) \quad (4)$$

3.3 The Case of Stockouts

Stockouts represent an important case of unobservable choice set heterogeneity, in which the demand model specifies the distribution of sales across the possible availability regimes (ie., the $y_{j,(a,x)}^{mis}$ terms). In this case, the change in choice set is endogenously determined. That is, a stockout occurs only when y_{kt} exceeds its capacity ω_{kt} . Consider a single stockout for the sake of exposition, with M_t consumers in a particular aggregate observation.⁹ One only needs to know how many of the M_t consumers saw choice set a_s and how many saw choice set a_t . Alternatively, one needs to know how many consumers were required in order to sell ω_{kt} units of product k under availability set a_s , since sales must be zero for k after the stockout. This defines a negative binomial distribution, which describes the number of failures r_t until ω_{kt} successes are observed. In the case of a stockout with periodic data, one also knows that the stockout happened before M_t consumers arrived. Thus, the stockout defines a negative binomial, conditional on the fact that $r_t + \omega_{kt} \leq M_t$.

The negative binomial distribution is:

$$f(r_t; \omega_{kt}, p_k) = \frac{(\omega_{kt} + r_t - 1)!}{r_t! (\omega_{kt} - 1)!} p_k^{\omega_{kt}} (1 - p_k)^{r_t}$$

And the conditional negative binomial:

⁹We provide the extension to the case of multiple stockouts in section A.2 of the appendix.

$$h(r_t; \omega_{kt}, p_k) \sim \frac{\text{NegBin}(\omega_{kt}, p_k)}{\text{NegBinCDF}(M_t - \omega_{kt}, \omega_{kt}, p_k)} = \frac{f(r_t, \omega_{kt}, p_k)}{\sum_{m=1}^M f(r_t, \omega_{kt}, p_k)}$$

Thus, $M_t - \omega_{kt}$ consumers bought some product other than the one that stocked out, and the number of those consumers in the market before the stockout occurred (r_t) has a conditional negative binomial distribution. Therefore, sales of product j that occurred prior to the stockout are distributed binomially. The expected sales of product j before and after the stockout of product k are given by:

$$\begin{aligned} E[y_{jt}^{a_s}] &= \sum_{r_t \leq M_t} y_{jt} \left(\frac{r_t \cdot p_j(a_s, x_t, \theta)}{r_t \cdot p_j(a_s, x_t, \theta) + (M_t - \omega_{kt} - r_t)} \right) h(r_t; \omega_{kt}, p_k) \\ E[y_{jt}^{a_t}] &= y_{jt} - y_{jt}^{a_s} \end{aligned}$$

This gives the direct ML estimator.

$$\begin{aligned} \hat{\theta}_{ML} = \arg \max & \\ & \sum_{j,t} \left[\sum_{r_t \leq M_t} y_{jt} \left(\frac{r_t \cdot p_j(a_s, x_t, \theta)}{r_t \cdot p_j(a_s, x_t, \theta) + (M_t - \omega_{kt} - r_t) \cdot p_j(a_t, x_t, \theta)} \right) h(r_t; \omega_{kt}, p_k) \right] \ln p_j(\theta, a_s, x_t) \\ & + \sum_{j,t} \left[\sum_{r_t \leq M_t} y_{jt} \left(\frac{(M_t - \omega_{kt} - r_t) \cdot p_j(a_s, x_t, \theta)}{r_t \cdot p_j(a_s, x_t, \theta) + (M_t - \omega_{kt} - r_t) \cdot p_j(a_t, x_t, \theta)} \right) h(r_t; \omega_{kt}, p_k) \right] \ln p_j(\theta, a_t, x_t) \end{aligned}$$

The challenge of estimation is that the sufficient statistics themselves now depend on the unknown parameters θ .¹⁰ However, there exists a trivial application of the EM algorithm. This is an iterative procedure in which every major iteration consists of two steps: the E-Step, in which the expected log likelihood $E[l(y_t|\theta^l)]$ is computed for the current iteration of θ^l ; and the M-Step, in which the expected likelihood is maximized over θ^{l+1} . This procedure increases the likelihood, $l(y_t|\theta^{l+1}) \geq l(y_t|\theta^l)$, at each iteration, as shown by Dempster, Laird, and Rubin (1977).

E-Step

Obtain a consistent expected value of the log-likelihood function at the current parameter estimates. In this case, it is sufficient to compute the expected value $E[y_t|\theta^l]$ at the current iteration of the parameters θ^l and plug it in.¹¹ In general, this need not be true for EM-type

¹⁰In addition, $\frac{\partial}{\partial \theta} E[y_t|a, x, \theta]$ may be complicated to obtain.

¹¹The multinomial is a special case of the EM algorithm dating back to Hartley (1958) where the log-likelihood is linear in the sufficient statistics and the expected log-likelihood is obtained by plugging in the imputed values. We omit the technical detail that allows us to work with expected log-likelihood instead of expected likelihood. This is the result of a trivial application of Jensen's inequality, well known in the EM literature (Dempster, Laird, and Rubin 1977).

procedures, but is a well-known feature of the multinomial distribution. In this case:

$$\begin{aligned}
E[l(y_t|\theta^l)] &= E[\sum_{jt} y_{jt} \ln p_j(\theta^l)] \\
&= \sum_{jt} E[y_{jt}|\theta^l] \ln p_j(\theta^l) \\
\hat{y}_{jt}^{a_s(k)} = E[y_{jt}^{a_s}|\theta^l] &= \sum_{r_t \leq M_t} y_{jt} \frac{r_t \cdot p_j(\theta^l)}{r_t \cdot p_j(\theta^l) + (M_t - \omega_{kt} - r_t) \cdot p_j(\theta^l)} h(r_t; \omega_{kt}, p_k(\theta^l))
\end{aligned}$$

M-Step

Maximize the expected log-likelihood function given the expected value $E(y_t|\theta^l)$:

$$\hat{\theta}^{l+1} = \arg \max_{\theta} \sum_{jt} \hat{y}_{jt}^{a_s(k)} \ln p_j(\theta, a_s, x_t) + \sum_{jt} \hat{y}_{jt}^{a_t(k)} \ln p_j(\theta, a_t, x_t)$$

Each iteration of the EM algorithm is guaranteed to improve the log-likelihood. When the stopping condition (i.e., $|\theta^{l+1} - \theta^l| < \varepsilon$) is met, the value of $\hat{\theta}$ represents a valid maximum of $E[l(y_t|\theta)]$.

There is no additional integration in the M-Step, so optimization routines treat the imputed values as if they were data. Thus, researchers can use ‘off-the-shelf’ routines for MSL. Furthermore, the expectations over the missing data do not enter the optimization procedure, thus avoiding computation of the derivatives. Similarly, one need not worry about the combinatorics of the problem growing, since one only evaluates the expectations once per major iteration, rather than at each likelihood evaluation.

In general there are some criticisms of EM-type procedures. The first is that EM algorithms are slow to converge (that is, they require many major iterations). This is true in that the rate is generally sub-linear for EM-type procedures. However, the rate depends on the amount of data that is missing, and how much the likelihood is affected by the missing data. For many cases involving stockouts, both of these tend to be small. That is, the bulk of the data is nonmissing (we don’t always observe stockouts), and for cases where we do observe stockouts, the choice probabilities before and after the stockout are generally not wildly different. The other problem that EM-type procedures generally have is that they are local optimizers (as are almost all nonlinear search routines used by economists) and require good starting values. This is true, and EM is a little more dependent on starting values than other algorithms, but this is a problem that is endemic to the extremum estimator class in general.

3.4 An Alternative Approach

Our approach is characterized by the following strategy. Given the conditional independence assumption, the occurrence of the stockout is distributed as a negative binomial with a fixed parameter. Conditional on stockout time, the expected sales of each product are binomially distributed. Expected sales can be obtained exactly by exhaustively integrating at every point in the domain of the stockout distribution. Finally, we plug these into the likelihood, which is linear in expected sales, and use an EM procedure for optimization. The primary advantage is that one only needs to consider potential orderings for the sales of products that experience a stockout, which reduces the combinatorial burden. We consider these orderings in an E-Step, which is done ‘offline’ from estimation.

One potential disadvantage of the approach just described is that it relies on an assumption of conditional independence. Stockouts imply a distribution on sales only when consumer preferences are the same before and after the stockout. As noted, this is actually a weaker requirement than the IID assumption used in most models of demand. However, either assumption may attract fresh scrutiny when applied to highly granular data. It is unclear whether conditional independence makes more sense over short periods of time (where we might expect to sample from different subgroups of the population at different times of day) or over longer periods of time (where we might expect to observe changes in the pool of consumers or their tastes). Our estimator leverages the conditional independence assumption because we use the fact that the distribution of tastes is fixed before and after a stockout event in order to assign sales availability to regimes. This in turn helps to define the distribution of sales across availability regimes separately from the θ vector.¹²

One might wonder whether alternative approaches exist that do not lean so heavily on the conditional independence of consumers. In general, these approaches require that one tracks the entire order of sales, rather than just the stockout time and expected sales, because different orderings can imply different likelihood values. One possibility is an indirect inference approach, where, for a guess of θ , one records the initial inventory and number of consumers for each market, and then simulates individual consumer choices, repeating for each of the consumers in the market.¹³ One would repeat this process some large number of times for each market and compute the average purchase probability for each product, denoting it as $\hat{p}_{jt}(\theta)$.¹⁴ If each consumer’s utility was determined by IID draws of random tastes and idiosyncratic preferences, then this would be a less efficient version of our estimator. On the other hand, the simulation approach allows the distribution of tastes to vary over time, even in a way that allows for correlation with stockouts. In fact, this indirect-inference type approach could be utilized for any scenario where one can write down a series of functions that allowed for simulation of consumer purchases. After computing the choice probabilities

¹²Without this assumption, one could still use a similar method, but the choice probabilities would follow more complicated mixture forms.

¹³Simulating the product choice for one consumer requires 4 steps: (1) draw a consumer type given θ , (2) compute choice probabilities given the type, (3) simulate a consumer from those choice probabilities and record a purchase, (4) update the inventory and number of consumers remaining in the market.

¹⁴Leslie (2004) employs a similar strategy for handling seat capacities in an analysis of theater demand.

from a simulation procedure, one would simply evaluate the likelihood using MSL.¹⁵

Even for an infinite number of trials, a simulation approach is not fully efficient like the EM procedure. The major efficiency loss arises because the procedure is not restricted to availability regimes that occur in the data. In other words, it does not condition on the full set of available data, and instead averages over all possible availability regimes encountered by simulation. Musalem, Olivares, Bradlow, Terwiesch, and Corsten (2008) show that it is possible to avoid this problem. They describe an MCMC procedure that assumes an ordering of sales, and then considers permutations of the ordering via Metropolis-Hastings steps. However, such an approach still requires considering (probabilistically) the full permutation of $M_t!$ possible orderings of the sales vector, whereas our approach only requires that we consider $\binom{M_t}{\omega_{kt}}$ orderings. Taking this together with the fact that the conditional independence assumption does not seem too far-fetched in our application, we do not report results from such a method here. However, we provide further details on this type of approach in the appendix (section A.3), and refer other researchers with different applications to that information.

4 Estimation

4.1 Parametrizations

All that remains is to specify a functional form for $p_j(\theta, a_t, x_t)$. In this section we present several familiar choices and how they can be adapted into our framework. In any discrete model, when n is large and p is small, the Poisson model becomes a good approximation for the sales process of any individual product. The simplest approach would be to parameterize $p_j(\cdot)$ in an semi-nonparametric way:

$$p_j(\theta, a_t, x_t) = \lambda_{j,a_t}$$

Then, the maximum likelihood (ML) estimate is essentially the mean conditional on (a_t, x_t) . This is more or less the approach that Anupindi, Dada, and Gupta (1998) take. The advantage is that it avoids placing strong parametric restrictions on substitution patterns, and the M-Step is easy. The disadvantage is that it requires estimating J additional parameters for each choice set a_t that is observed. It also means that forecasting is difficult for a_t 's that are not observed in the data or are rarely observed, which highlights issues of identification that we will address later. Furthermore, the lack of a utility-based framework means that out-of-sample predictions about alternative policies cannot be made.

A typical solution from the differentiated products literature for handling these sorts of problems is to write down a nested logit or random coefficients logit form for choice probabilities. This still has considerable flexibility for representing substitution patterns,

¹⁵One could also use an MSM procedure, as in Berry, Levinsohn, and Pakes (1995), to allow for price instruments or additional moment restrictions.

but avoids estimating an unrestricted covariance matrix. This family of models is also consistent with random utility maximization (RUM). If we assume that consumer i has the following utility for product j in market t and they choose a product to solve:

$$\begin{aligned} d_{ijt} &= \arg \max_j u_{ijt}(\theta) \\ u_{ijt}(\theta) &= \delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2) + \varepsilon_{ijt} \end{aligned}$$

where δ_{jt} is the mean utility for product j in market t , μ_{ijt} is the individual specific taste, and ε_{ijt} is the idiosyncratic logit error. It is standard to partition the parameter space $\theta = [\theta_1, \theta_2]$ between the linear (mean utility) and non-linear (random taste) parameters. This specification produces the individual choice probability, and the aggregate choice probability

$$Pr(k|\theta, a_t, x_t) = \frac{\exp[\delta_k(\theta_1) + \mu_{ik}(\theta_2)]}{1 + \sum_{j \in a_t} \exp[\delta_j(\theta_1) + \mu_{ij}(\theta_2)]}$$

This is exactly the differentiated products structure found in many IO models (Berry 1994, Berry, Levinsohn, and Pakes 1995, Goldberg 1995). These models have some very nice properties. The first is that any RUM can be approximated arbitrarily well by this “logit” form (McFadden and Train 2000). This also means that the logit ($\mu_{ijt} = 0$) and the nested logit can be nested in the above framework. For the random coefficients logit of Berry, Levinsohn, and Pakes (1995), $\mu_{ijt} = \sum_l \sigma_l x_{jl} \nu_{il}$, where x_{jl} represents the l th characteristic of product j and ν is standard normal.

The second advantage of these parametrizations is that it is easy to predict choice probabilities as the set of available products changes. If a product stocks out, we simply adjust the a_t in the denominator and recompute. A similar technique was used by Berry, Levinsohn, and Pakes (1995) to predict the effects of closing the Oldsmobile division and by Petrin (2002) to predict the effects of introducing the minivan. The parsimonious way of addressing changing choice sets is one of the primary advantages of these sorts of parameterizations, particularly in the investigation of stockouts.

We estimate the model using two specifications for the choice probabilities: a nested-logit model (with either a single nesting parameter or category-specific nesting parameters), and a random-coefficients model. For the nested-logit model, we denote product-specific effects as d_j , a nesting parameter as λ . We use Full Information Maximum Likelihood, and reparameterize the model such that $\gamma_j = d_j/\lambda$. This gives the choice probabilities as

$$\hat{p}_{jt}(\theta) = \frac{e^{\gamma_j} (\sum_{k \in g_l} e^{\gamma_k})^{\lambda-1}}{\sum_{\forall l'} (\sum_{k \in g_l'} e^{\gamma_k})^{\lambda}}$$

For the random-coefficients logit form, we estimate via Maximum Simulated Likelihood

(MSL). For the multinomial choice model, the MSL estimator begins with some random or quasi-random normal draws v_{ik} for each t in the dataset. For a given θ we can compute the average choice probability across draws and then plug this in to our likelihood function. The average choice probabilities are given by:

$$\hat{p}_{jt}(\theta) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp[d_j + \sum_l \sigma_l v_{il} x_{jl}]}{1 + \sum_{j \in a_t} \exp[d_j + \sum_l \sigma_l v_{il} x_{jl}]} \text{ where } v_{il} \sim N(0, 1).$$

Finally, for either specification, the likelihood is simply:

$$l(\theta) = \sum_t \sum_{j \in a_t} y_{jt} \ln \hat{p}_{jt}(\theta).$$

4.2 Heterogeneity

Thus far, choice probabilities have been conditional on x_t . This is useful for showing that our result holds for the case of conditional likelihood, but it is also of practical significance to applied problems. It is likely that choice probabilities may vary substantially across periods in retail datasets, particularly if they are short. Over long periods of time (such as annual aggregate data) these variations get averaged out. The distribution of tastes over a long period is essentially the combination of many short-term taste distributions, and this is often the basis of estimation (ie., in the case of limited data we would estimate the long run distribution). With high frequency data, we are no longer so limited and can address this additional heterogeneity by conditioning on x_t , which could include information such as the time of day, day of the week, or local market identifiers. Depending on how finely the data are observed, not accounting for this additional heterogeneity may place a priori unreasonable restrictions on the data.

We can model dependence on x_t in several ways. One is to treat $p(\cdot|x_t)$ as a different function for each x_t . Another is to require that all markets face the same distribution of consumers, but allow that distribution to vary with x_t . Another option is to fix some parameters across x_t , and allow others to vary with x_t . For example, allow mean tastes to differ across markets but assume that the correlation of tastes is constant. In addition, one can parameterize market size, M , as a function of x_t (i.e., allow market size to vary across periods without affecting the choice probabilities). Parameterizing M with auxiliary data has a long history in the literature (Berry 1992), and is done offline prior to parameter estimation.

4.3 Identification

In this section we address practical aspects of identification of the choice probabilities $p_j(\theta, a_t, x_t)$, when the underlying data generating process is multinomial. Denote $y_j(a, x)$ as sales of product j when $(a_t = a, x_t = x)$, and $M_{(a,x)}$ as the corresponding market size.

Since sales are distributed multinomially, the semi-nonparametric estimator for \hat{p}_j is just the conditional mean, or the fraction of consumers facing (a, x) who chose product j . This is:

$$\hat{p}_j(a, x) = \frac{y_j(a, x)}{M_{(a, x)}}$$

with variance:

$$\begin{aligned} \text{Var}(\hat{y}_j(a, x)) &= M_{(a, x)} p_{j(a, x)} (1 - p_{j(a, x)}) \\ \text{Var}(\hat{p}_j(a, x)) &= \frac{1}{M_{(a, x)}^2} \cdot \text{Var}(y_{j(a, x)}) = \frac{p_j(a, x)(1 - p_j(a, x))}{M_{(a, x)}} \end{aligned}$$

Thus, the variance of nonparametric estimators for $p_j(a, x)$ will go to zero as $M_{a, x} \rightarrow \infty$. This variance is typically referred to as “measurement error in the choice probabilities,”¹⁶ and it declines almost uniformly across share sizes.¹⁷

Conversely, unless every (a, x) pair in the domain is observed with a substantial number of consumers, the semi-nonparametric representation of $p_j(a, x)$ will not be nonparametrically identified. This problem is well known, and the standard approach is to assume a functional form for the $p_j(\cdot)$ ’s, which may also specify an underlying utility function. One then determines what variation is needed in the data to identify this function. There is a growing literature on identification results for this class of estimators starting with Matzkin (1992). For example, Akerberg and Rysman (2005) use continuous variation in product characteristics (such as price) to obtain some derivative-based identification arguments for choice probabilities in the nested logit. More recent work by Berry and Haile (2008) and Fox and Gandhi (2008) provide formal identification results for the latent utilities using continuous full-support variation in product characteristics similar to the special regressor econometric literature.

Our approach is different, because it is not motivated by continuous full support x variation. Instead stock-outs or capacity constraints generate quasi-experimental variation among available choice sets. We use this short-run variation to identify the distribution of tastes and substitution across products. The underlying intuition is simple: The easiest way to identify substitution patterns is to exogenously remove products from the choice set, and record the sales.

If we observe only a single choice set, then we have only a single observation in our data. However, we could identify product dummies for every product that was available for the logit form from this single observation (in the absence of any random coefficients).

¹⁶For ML estimators, “measurement error” is an efficiency issue. However, for GMM approaches, this can create problems with consistency as well, so we typically assume that $M_{a, x} \rightarrow \infty$. While this might be reasonable for annual data at a national level, it becomes more problematic in the analysis of high-frequency data.

¹⁷Thus we don’t need more observations to correctly estimate smaller shares than we need to estimate larger shares.

Identification of nonlinear parameters in a random-coefficients specification comes from the fact that the sales of two products j and k will be differentially affected by a stockout of product l depending on how similar x_j and x_k are to x_l . Specifically, for each consumer type i we inflate the probability of buying good j by a factor proportional to the probability that type i bought the stocked-out good l . Thus, we can think about a stockout as providing information not only about the level of p_{jt} , but also the ratio of the choice probabilities before and after a stockout.¹⁸

Intuitively, one can think of markets as being of three types. In type 1, no products stock out. This is the usual scenario. In type 2, a product is stocked out at the very beginning of the period. This creates good variation in the data—we have exogenous short-run variation in choice sets, which is not chosen by the firm. We need only adjust the set of available products accordingly in estimation. Having perpetual data collection will guarantee that we always observe either type 1 or type 2 observations. In type 3 markets, a product stocks out sometime between periodic data collection points. Periodic data collection will generate these types of observations in addition to types 1 and 2. Ignoring these observations selects on sales levels because higher than average demand periods are more likely to experience stockouts, and so one needs to incorporate these periods, which is exactly what our missing data procedure does.

When type 3 markets are observed in a dataset, we need to identify both the distribution of the latent availability regimes and the choice probabilities. As discussed in the model section, the distribution of the latent availability regimes is specified by the multinomial model of sales. Separate identification of these two components (the choice probabilities and the distribution of availability regimes) arises from two sources of variation: first, variation in choice sets across all market types (including type 1 and type 2 markets), and second, variation in starting inventory levels in type 3 markets. The latter variation helps to pin down the distribution of latent availability regimes. In principle, one could identify the model from either source of variation, although the presence of type 1 and type 2 markets is clearly helpful. If working with data from only type 3 markets, one would want extensive variation in starting inventory levels in order to separately identify the two components of the model.

In practice, much empirical work uses a characteristics-based approach, and relies on discrete changes in stocking decisions, or non-price product characteristics, and a finite amount of (possibly small) price variation. As Akerberg and Rysman (2005) point out, it’s important to observe variation in price, and not just discrete or non-price characteristics, in order to identify price sensitivity. And, furthermore, small price variation across a tight range will generally provide less information than well-spaced price variation. In our vending application, we don’t observe price variation, and we do not attempt to identify a price coefficient. In general, one would want variation across prices in addition to changes in the choice set in order to reliably estimate a price coefficient.¹⁹

¹⁸We provide this ratio and its derivation in section A.4 of the appendix.

¹⁹Our estimation method allows for inclusion of a price instrument by either the addition of distributional assumptions to the ML problem, or the use of a GMM procedure. Draganska and Jain (2004) develops a

In summary, our model presents a different way to interpret variation in choice sets, in which availability varies as products stock-out, rather than with potentially endogenous changes in the long-term product mix or pricing decisions. This variation is exogenous because choice sets are realizations of the stochastic choices of consumers, and as a result, firms and consumers both take these short-run changes in choice sets as given. While firms choose inventory or capacity to affect the probability that a product stocks out, once we condition on that probability, the occurrence of any particular stockout is random.

5 Industry Description, Data, and Reduced-form Results

5.1 The Vending Industry

The vending industry is well suited to studying the effects of product availability in many respects. Product availability is well defined: goods are either in-stock or not (there are no extra candy bars in the back, on the wrong shelf, or in some other customer’s hands). Likewise, products are on a mostly equal footing (no special displays, promotions, etc.). The product mix, and layout of machines is relatively uniform across the machines in our sample, and for the most part remains constant over time. Thus most of the variation in the choice set comes from stockouts, which are a result of stochastic consumer demand rather than the possibly endogenous firm decisions to set prices and introduce new brands.²⁰

Typically, a location seeking vending service requests sealed bids from several vending companies for contracts that apply for several years. The bids often take the form of a two-part tariff, which is comprised of a lump-sum transfer and a commission paid to the owner of the property on which the vending machine is located. A typical commission ranges from 10 – 25% of gross sales. Delivery, installation, and refilling of the machines are the responsibility of the vending company. The vending company chooses the interval at which to service and restock the machine, and collects cash at that interval. The vending company is also responsible for any repairs or damage to the machines. The vending client will often specify the number and location of machines.

Vending operators may own several “routes” each administered by a driver. Drivers are often paid partly on commission so that they maintain, clean, and repair machines as necessary. Drivers often have a thousand dollars worth of product on their truck, and a few thousand dollars in coins and small bills by the end of the day. These issues have motivated advances in data collection, which enable operators to not only monitor their employees,

method for including IVs and supply-side restrictions into an ML estimator, assuming a normal distribution on the unobservable product attributes. For a GMM procedure, Berry (1994) and Berry, Levinsohn, and Pakes (1995) can be used in the M step, because these estimators improve the likelihood at each step. However, these methods rely on an assumption that individual markets are large ($M_t \rightarrow \infty$), which might be problematic in very granular datasets. For the most ‘extreme case’ of granularity, Chintagunta, Dube, and Goh (2005) provide a method to extend the Berry (1994) method with price IVs to cases with individual-level data. Finally, Allenby, Chen, and Yang (2003) provide a Bayesian estimator when IVs are required.

²⁰In this sense, our setup is substantially simpler than that of Nevo (2001), Goldberg (1995), or Berry, Levinsohn, and Pakes (1995) where new brands and prices are substantial sources of identification.

but also to transparently provide commissions to their clients and make better restocking decisions.

Machines typically collect internal data on sales. The vending industry standard data format (called Digital Exchange or DEX) was originally developed for handheld devices in the early 1990's. In a DEX dataset, the machine records the number and price of all of the products vended; these data are typically transferred to a hand-held device by the route driver while he services and restocks the machine. The hand-held device is then synchronized with a computer at the end of each day.

5.2 Data Description

In order to measure the effects of stock-outs, we use data from 54 vending machines on the campus of Arizona State University (ASU). This is a proprietary dataset acquired from North County Vending with the help of Audit Systems Corp (later InOne Technologies, now Streamware Inc.). The data were collected from the spring semester of 2003 and the spring semester of 2004. The ASU route was one of the first vending routes to be fully wireless enabled and monitored through Audit System's (now Streamware's) software. The wireless technology provides additional inventory observations between service visits, when the DEX data are wirelessly transmitted several times each day (approximately every four hours).

The dataset covers snack and coffee machines; we focus on the snack machines in this study. Throughout the period of observation, the machines stock chips, crackers, candy bars, baked goods, gum/mints, and a few additional products. Some products are present only for a few weeks, or only in a few machines. Of these products, some of them are non-food items²¹ or have insubstantial sales (usually less than a dozen total over all machines).²² In our analysis, we exclude these items in addition to gum/mints, based on an assumption that these products are substantially different from more typical snack foods (and rarely experience stockouts). For a few brands of chips, we observe rotation over time in the same slot of the machine, and for these goods, we create two composite chip products (Misc Chips 1 and Misc Chips 2).²³ Finally, we combined two different versions of three products.²⁴ The 44 products in the final dataset are listed in Table 1.

Sales at each machine are observed at four-hour intervals. Retail prices are constant over time, machines, and across broad groups of products as shown in Table 1. Baked goods typically vend for \$1.00, chips for \$0.90, cookies for \$0.75, and candy bars for \$0.65. As

²¹While often sold alongside of snacks in vending machines, condoms are poor substitutes for potato chips.

²²Products dropped for insubstantial sales are: Grandma's Lemon Cheese, Grandma's Chocolate Croissant, and Nestle 100 Grand.

²³Misc Chips 1 rotates: Cool Ranch, Lays Kettle Jalapeno, Ruffles Baked Cheddar, and Salsa Dorito. Misc Chips 2 rotates: Frito Jalapeno, KC Masterpiece BBQ, Lays Baked Potato, Lays Wisconsin Cheese, Rubbles Hearty Chili, and Frito Chili Cheese. Product characteristics for the goods that are combined are very similar; for the composite good, we use the average of the characteristics of the individual products.

²⁴These were: combine Gardetto's with Gardettos Snackems, combine Nestle Crunch with Caramel Nestle Crunch, and combine Nutter Butter with Nutter Butter Bites. Product characteristics in the first two combinations are identical. In the last combination, the product characteristics differ slightly, and in that case, we use the characteristics from Nutter Butter Bites.

compared to typical studies of retail demand and inventories (which often utilize supermarket scanner data), there are no promotions or dynamic price changes as in Nevo and Hendel (2007) or Aguirregabiria (1999). This means that once most product characteristics (and certainly product or category dummies) are included, price effects are not identified. The method we present will work fine in cases where a price coefficient is identified, but in our particular empirical example, we have no variation for identifying a price effect.

In addition to the sales, prices, and inventory of each product, we also observe product names, which we link to the nutritional information for each product in the dataset. For products with more than one serving per bag, the characteristics correspond to the entire contents of the bag.

The dataset also contains stockout information and marginal cost data (the wholesale price paid by the firm) for each product. The stockout percentage is the percentage of time in which a product is observed to have stocked-out. We report both an upper and a lower bound for this estimate. The lower bound assumes that the product stocked out at the very end of the 4-hour period we observe, and the upper bound assumes that it stocked out at the very beginning of the 4-hour period of observation. For most categories and products, this ranges from two to three percent, with larger rates of stockouts for pastry items. The marginal cost data are consistent with available wholesale prices for the region. There is slight variation in the marginal costs of certain products, which may correspond to infrequent re-pricing by the wholesaler. The median wholesale prices for each product are listed in Table 1. Table 1 also allows one to calculate markups of the products. Markups tend to be lowest on branded candy bars (about 50%), and high on chips (about 70%). The product with the highest markup is Peter Pan crackers, which has a markup of 84%.

Other costs of holding inventory are also observed in the data, including spoilage/expiration and removal from machines for other reasons (e.g., ripped packaging, contamination, etc.). Spoilage does not constitute more than 3% of most products sold. The notable exceptions are the Hostess products, which are baked goods and have a shorter shelf life than most products (approximately 2 weeks vs. several months). For our static analysis of demand, we assume that the costs associated with such events are negligible.

5.3 Reduced-form Results

Before applying the estimation procedure described above to the dataset, we first describe the results of two simple reduced-form analyses of stockouts. In table 2, we compute the profits for each four-hour wireless time period at each individual machine and regress this on the number of products stocked out. The first specification (Column 1) estimates the four-hour profit loss to be about \$0.90 per product stocked out. Column 2 allows the effect of a stockout to differ based on the number of stockouts in the category with the most stockouts, in order to capture the fact that substitution to the outside good may increase when multiple products are unavailable in the same category (ie., missing one candy bar and one brand of chips is different from missing two brands of chips). We estimate the effect of a stockout in the category with the most products missing to be about \$1.90 per four-hour

period, and the base effect of a product stocking out to be \$0.44. In column 3, we include the number of stockouts in each separate category. These results estimate the costs per stockout at around \$1.45 for chips to \$3.85 for candy on top of a base effect of \$0.41. The results are robust to the inclusion of machine fixed effects, which explain an additional 20 percent of the variation in profits. All of these regressions are clearly endogenous, and may be picking up many other factors, but they suggest some empirical trends that can be explained by the full model. Namely, stockouts decrease hourly profits as consumers substitute to the outside good, and multiple stockouts among similar products causes consumers to substitute to the outside good at an increasing rate.

Table 3 reports the results of a regression of stockout rates on starting inventory levels. An observation is a service visit-product pair. We report results for Probit and OLS (Linear Probability Model) with and without product and machine fixed effects. We find that an additional unit of inventory at the beginning of a service period reduces the chance of a stockout in that product by about 1%. A full column of candy bars usually contains 20 units. This means that the OLS (fixed effects) probability of witnessing a stockout from a full machine in a 3-day period is $.238 - .0101 * 20 = 3.6\%$. For a product at a machine with a starting inventory of five units, the predicted chance of a stockout is about one in five.

6 Empirical Results

We estimate nested logit and random-coefficients logit demand specifications using three different treatments for stock-out events. In the first treatment we assume full availability in all periods, including those periods in which a stockout was observed. Choice set variation in this specification is generated by the introduction or removal of products over time, and, to a lesser extent, from selective stocking of products in different machines. We refer to this as the ‘Full Availability’ model, and it is the standard method of estimation in the literature. In the second treatment we account for stock-outs that were fully observed, but ignore data that were generated during periods in which the timing of a stockout was ambiguous. We call this the ‘Ignore’ model. In the third treatment, we account for fully-observed stock-out events, and use the EM algorithm to estimate which sales occurred under the various stock-out regimes within any ambiguous period.²⁵ This is the ‘EM-corrected’ model.

Overall sales levels in the data vary across vending machines and time periods (such as time-of-day or day-of-week indicators), but relative choice probabilities are remarkably

²⁵The data contain a small number of observations (less than one percent) in which three or more products stock out. Based on conversations with the vendor, we assume such events occur at the very end of any ambiguous period for the estimates of demand reported here. The results are robust to omitting these observations, which the vendor believed may contain coding errors, or indicate removal or replacement of a machine. While inclusion of these observations—were we to believe the data from them fully—is possible for estimation, the simpler treatment of them here eases the computational burden in our application. For settings in which large numbers of products stock out within a period of observation, refer again to the methods in section A.3 of the appendix on alternative computational methods, which avoid integration of the exact distribution.

similar, and so we accommodate heterogeneity through M .²⁶ We divide machines into three tiers based on overall sales levels, and multiply a base rate of 360 people per day per machine by one-third and 3 for the smallest and largest tiers respectively. In addition, the rate of arrivals is reduced by a factor of one-third at night and one-fifth on weekends. Thus, for a machine in the middle tier, we have 250 consumers in a weekday, and about 75 consumers on a weekend day, for a total of 1400 consumers per week on average. For machines in the lowest tier, this is roughly 467 consumers per week on average, and for a machine in the highest tier, this is about 4200 consumers per week on average.²⁷ Due to data limitations, we do not allow choice probabilities to vary across locations or time periods. The most important limitation for the data is that we want to be sure there are enough potential consumers in any particular availability set. The more conditioning we do, the fewer consumers we have for any given availability set. For example, if we allowed choice probabilities to differ freely for each machine, we could only use the observations generated by that machine for each availability set that that machine encountered.

In addition to the results reported here, which use the complete dataset, we also estimated the model after aggregating the data to the daily level. We did this to insure that the EM correction does not perform poorly when ‘ambiguous’ periods comprise a larger portion of our dataset. At the four-hourly level, approximately 17 percent of the data come from periods during which a stock-out occurred, versus 35 percent when the data are aggregated at the daily level. We estimate very similar results for both the disaggregated and aggregated data.²⁸

Table 4 reports the number of choice sets, market size, log likelihood, and nonlinear parameters from estimation of nested-logit and random-coefficients specifications under each of the three treatments of stock-out events.²⁹ The first two panels report estimates from two nested-logit specifications, the first using a single nesting parameter, and the second using five category-specific nesting parameters. Both nested-logit models are estimated by full information maximum likelihood methods. We report the nesting parameter λ from McFadden (1978) rather than the σ correlation parameter from Cardell (1997) or Berry (1994).³⁰ Roughly speaking, $\lambda \approx (1 - \sigma)$ such that $\lambda = 1$ is the simple logit and $\lambda = 0$ is perfect correlation within group. In general a $\lambda > 1$ is allowed, but is not necessarily consistent with random utility maximization (McFadden and Train 2000).

²⁶Thus, it is assumed that the choice probabilities (the $p_j(\cdot)$ functions) are stable across markets as discussed in the estimation section.

²⁷We use some additional conditions to prevent sales from exceeding the market size, particularly in very short periods, but in general these are not binding.

²⁸All daily-level results are available upon request from the authors.

²⁹All results were obtained by using the KNITRO optimization package. All reported values satisfy first and second order conditions for valid optima. A number of different starting values were used in estimation, and the best optimum value was reported in each case. These results have also been checked against standard MATLAB packages (fminsearch, fminunc).

³⁰We should reiterate the reason we use FIML rather than the simpler least squares estimator for the nested logit is that we worry not only about the endogeneity of $\ln(s_{j|g})$ the within group share, and the lack of potential instruments, but for many small markets the only within-group sales are the sales of product j – the case of extreme measurement error.

In the first column of the first panel (under the ‘Full Availability’ model), 238 choice sets contribute to the estimation of the ML problem, and total market size is about 5.7 million consumer visits over the two semesters. The correlation within nest is approximately $(1 - \lambda)$, or 0.48 under an assumption of full availability. The second column (the ‘Ignore’ model) uses 2649 unique choice sets and a market size of about 5.3 million consumer visits. The estimate of λ in this specification, 1.08, is not consistent with utility maximization, highlighting the extent of the bias from ignoring periods in which stock-outs occurred. The likelihood improves, but is not directly comparable to the likelihood in the other columns because it applies only to a subset of the data. The last column (‘EM’) reports the results after using the EM correction to assign sales to different availability regimes in the ambiguous periods. The number of choice sets in this model is 3966, which incorporates probabilistic choice sets, as well as those that were only encountered as an intermediary between the beginning-of-period and end-of-period availability in the case of multiple stockouts. The EM-corrected estimate of λ , 0.77, implies a within-nest correlation of 0.23. The EM-corrected model has a superior log-likelihood to the Full Availability model, although due to the biased estimates of θ in the Full Availability model, the likelihoods of the two models are not directly comparable.³¹

Standard errors are provided for all estimates. Standard errors for the Full Availability and Ignore models are readily available. The EM-corrected model requires a correction to the usual standard errors to account for the fact that sales in periods of latent availability regimes are estimated. We provide this correction in section A.6 of the appendix. In general, all parameters are estimated fairly precisely, with significant differences between the three models of stock-outs. The EM-corrected model produces the most efficient results by incorporating many more choice sets in estimation (in spite of the correction to account for estimating sales during periods of latent availability).

The second panel of table 4 reports similar patterns in the nested-logit model with five nesting parameters. In this specification, the estimated correlation between products in the same nest is negative for some nests in both the Full Availability and Ignore models, again highlighting the bias from ignoring stock-outs or dropping periods in which they occur. The EM-corrected model shows sensible correlation patterns, with within-nest correlation highest for chocolate $(1 - \lambda) = 0.54$, and lowest for pastry $(1 - \lambda) = 0.17$.

The third panel of table 4 reports the same set of estimates using a random-coefficients

³¹One could also imagine estimating a model in which sales during periods of unknown availability are arbitrarily assigned—for example, by assuming that all stocked-out products stock out either at the very end or the very beginning of any ambiguous period. One might expect that the likelihood from such an exercise would be improved by the application of the EM algorithm, and in that sense, provide a further check of the EM-corrected method used here. While this is true for cases in which the unobserved data do not depend on \mathbf{y} (see Tanner and Wong (1987)), it need not hold in the case of stock-outs, where the missing data include sales. Put another way, consistent estimation of demand implies a distribution for the missing sales, as detailed in section 3. Substituting arbitrary distributions instead will not give consistent estimates of θ , because the demand model implies a specific distribution as a function of observed sales. A similar argument applies for comparing the likelihood from the Full Availability case with the likelihood from the EM-corrected estimator, as these are not comparable for the same reason. Essentially, any such exercise injects data that are known to be false into the estimate of θ , making the resulting likelihood incomparable to the likelihood for the true model. We provide additional detail on this point in section A.5 of the appendix.

logit model for demand in which random coefficients are estimated for each of three observable product characteristics: fat, salt, and sugar.³² The random-coefficients specification has a higher likelihood than either of the nested-logit specifications under the ‘Full Availability’ assumption. However, the likelihood under both of the nested-logit specifications exceeds the random-coefficients specification under the EM-corrected model. In the case of Ignore, the random-coefficients specification favors no correlation in taste for the three observable characteristics (i.e., estimation reduces to a ‘plain vanilla’ logit model).³³

Using the EM-corrected estimates from table 4, table 5 lists the best substitute for the 35 most commonly-held products according to the category-specific nested-logit and the random-coefficients specifications. In both cases, we get sensible substitution patterns, with the two specifications predicting the same best substitute in many cases (particularly among chips and chocolate bars). When the predicted best substitutes differ, we see the trade-offs between the two demand specifications. For products that are ‘harder to categorize,’ such as Oreos (in the candy category), the random-coefficients model gives more intuitive results, whereas for products with less helpful characteristics, such as PopTarts, the nested-logit model seems more sensible.

7 Estimated Sales and the Impact of Stockouts

In this section, we use the results from the three estimated models to predict sales and the impact of stock-out events on firm profitability. These predictions give an indication of how important the corrections to the demand system are likely to be. They also lie at the heart of supply-side decisions about capacity and restocking efforts, and play a fundamental role in determining welfare calculations on the impacts of mergers, the value of new products, or the application of antitrust policy. Table 6 shows predicted weekly sales for a fully-stocked ‘typical’ machine under the category-specific nested-logit and random-coefficients specifications. A typical machine is defined as one carrying the 35 most widely-carried products (measured across machines and over time) out of the full set of 44 products in the data, for which we simulate the arrival of 4500 consumers.³⁴ Comparing Ignore to Full Availability for this typical machine shows that predicted sales levels under Ignore are substantially lower for all products except three (Chocolate Donuts, Ding Dong, and PopTart—under random coefficients). This highlights the bias that results from excluding

³²We have one additional product characteristic that is continuous: calories. We found no effect on correlation in tastes from this variable, so it was excluded from the set of non-linear parameters. We observe two additional discrete product characteristics: cheese and chocolate dummies. These are excluded from estimation because they were not identified after the inclusion of product dummies. We believe the non-identification of these parameters in our particular setting is due to the lack of additional product characteristics that vary continuously, such as price. Such a characteristic is a key assumption more generally for identification (see Berry and Haile (2008)).

³³We report estimates of the linear parameters (i.e., the product dummies) and the results of second-stage regressions of product dummies on characteristics in section A.7 of the appendix.

³⁴We chose 35 products because this is the number of product facings in a single machine. The market size of 4500 consumers is the number of consumers assumed to pass by a relatively high-volume machine over the course of one week in our demand model.

data on periods in which sales exceed inventory. The most interesting comparison is between the Full Availability and EM-corrected models, and we report the difference between these models in the fourth and eighth columns of the table. Looking across categories, pastry and chip products generally have higher sales under the EM-corrected model. These are the two categories with the lowest capacities and highest average rates of stock-outs in the data (see table 1). Within category, we also see some sensible patterns. For example, among chocolate bars, three products have higher estimated sales under the EM-corrected model: Snickers, M&M Peanut, and M&M’s. These three products have the highest rate of stock-out events in the chocolate category in the data (not including Baby Ruth, which was not carried by the ‘typical’ machine, and so was excluded from the simulation). Overall, the EM-corrected model predicts more people purchasing an inside good (total sales of 250, compared to 245 or 229 under Full Availability or Ignore).

In order to demonstrate how the different demand estimates affect the impact of stock outs on profitability, we conduct an experiment in which we consider weekly sales at the same typical machine, and compare them to a machine where the two best-selling products in each category are unavailable.³⁵ We report the results of these stockout experiments in Table 7. For each of the products stocked out, we report the number of forgone sales predicted by the two demand specifications. This ranges from roughly 10 in the pastry and chips categories, 6 or 7 in the cookie and candy categories, and 20 in the chocolate category.³⁶

The different treatments of stock-out events give substantially different predictions for sales as the set of available products changes. For both demand specifications, the Ignore model predicts much lower sales of available products across the board than either the Full Availability or the EM-corrected models. This reflects lower estimated correlation in tastes (the exception being the chocolate category in the nested-logit specification), but also highlights the bias that results from dropping periods of high demand from estimation.

In the case of the category-specific nested logit, the λ ’s that are greater than one for the Full Availability and Ignore models (i.e., the pastry category in both models and the chips category in the Ignore model) lead to the prediction that fewer consumers buy other products in the category when the top-selling products are stocked out. Thus, for example, a PopTart appears to be a complement to Ding Dongs and other pastry items in these specifications. For the remaining categories, the EM-corrected model generally predicts lower correlation in tastes (higher λ ’s), and thus fewer additional sales for the available substitutes (e.g., Peanuts sells 2.31 additional units according to the Full Availability model, but only 1.81 additional units under the EM-corrected model). The exception is the chocolate category, in which the EM-corrected model predicts higher within-nest correlation of tastes (a lower λ), and correspondingly higher rates of substitution to the available substitutes within the nest (e.g.,

³⁵We simulate the removal of the following products: Chocolate Donuts, Strawberry Frosted PopTarts, Grandma’s Oatmeal Raisin Cookie, Chips Ahoy Cookies, Rold Gold Pretzels, Sunchips Harvest Cheddar, Snickers, Twix, Starburst, and Kar Nut’s Sweet & Salty Mix.

³⁶As an interesting supply-side comparison, the forgone sales of these products match up in a sensible way against the capacity of the machine, which is visited once a week or slightly more often. Capacities in the various categories are: 9 to 11 for pastry and most chips, 15 for cookie and candy, and 20 for most chocolate bars.

5.16 units of M&M Peanut are sold according to the EM-corrected model, compared to 2.96 units under the Full Availability model).

In the random-coefficients specification, the EM-corrected estimates predict less substitution to available products than the Full Availability model across the board. This demonstrates the censoring and forced-substitution effects, because the products that were stocked-out are the best-selling products in each category, and they also stock-out more frequently than other products. (One exception to this rule is Twix, which suffers fewer stock-outs in the data.) The EM-corrected model, under both demand specifications, generally predicts demand that is stronger for the best-selling, frequently stocked-out products (giving larger negative numbers of forgone sales for those products), and weaker for the remaining available products.

In table 8, we report the overall impact of the stockouts. The Full Availability and Ignore models predict lower levels of forgone sales compared to the EM-corrected model under both demand specifications. The Full Availability model predicts higher increased sales of substitutes than the EM-corrected results under the random-coefficients specification, while the Ignore model predicts ridiculously low levels of substitution to the available products. Correspondingly, the Ignore model predicts a much lower percentage of consumers purchasing an inside good when their first-choice product is unavailable (between 3 and 9 percent), while the Full Availability and EM-corrected models are closer to each other, with the EM-corrected model predicting fewer consumers staying inside (25 - 36 percent vs. 30 - 35 percent for Full Availability). Gross profit is thus lower under the EM-corrected model, with a loss from the stockouts of roughly \$36 to \$40 compared to \$33 to \$36 for the Full Availability model. In percentage terms, this difference is 7.9% in the nested-logit specification, and 11.6% in the random-coefficients specification. However, the predictions in each category are wildly off, with the change in gross profit differing by between 7% to 30% in the random-coefficient specification, and by even more in the nested-logit specification. For example, the nested logit model predicts only a \$2.41 loss in profits in the Chips category in the Full Availability model, vs. a \$10.32 loss in the EM-corrected results. For an industry with profit margins of less than 4%, in which category-level sales are important, these are significant differences.³⁷

As a final exercise, we quantify the expected change in sales for substitute products, given that a focal product stocks out. Doing this for each of the products stocked in a typical machine produces 35 graphs with 35 sales effects in each one. The first bar in these graphs is the closest substitute, the second bar is the next closest substitute, and so on. Figure 8 shows the median change in the sales of substitutes by rank across all 35 products using the demand parameters from the EM-corrected random-coefficients model.³⁸ The median effect shows that the closest substitute experiences about a three percent increase in its sales when the focal product stocks out. The second closest substitute has a median sales increase of two percent.

³⁷Companies with over \$1 million in revenue have a 4.3% profit margin on average, while companies with less than \$1 million in revenue (75% of all vending operators, by count) have an average profit margin of -2.5% (www.vending.org 2008).

³⁸The set of 35 individual graphs are available upon request from the authors.

8 Conclusion

Incomplete product availability is a common and important feature of many markets. This paper demonstrates that failing to account for product availability correctly can lead to biased estimates of demand, which can give misleading estimates of sales and the welfare impacts of stock outs. We show that the welfare impact of stockouts in vending machines has a substantial effect on firm profits, indicating that product availability may be an important strategic and operational concern facing firms and driving investment decisions. Furthermore, biases that result from the incorrect treatment of stock-out events can potentially undermine the reliability of many important applications of demand estimates for markets with incomplete product availability, such as simulating the welfare implications of mergers or new product introductions, applying antitrust policy, constructing price indices, or evaluating the optimal capacity choices of firms.

A failure to account for product availability also ignores a useful source of variation for identifying demand parameters. Rather than examining the effect of changing market structure (entry, exit, new goods, mergers, etc.) on market equilibrium outcomes, stock outs allow us to examine the effect that temporary changes to the consumer's choice set have on producer profits and our estimators. Standard demand estimation techniques have used long-term variations in the choice set as an important source of identification for substitution patterns, and this paper demonstrates that it is also possible to incorporate data from short-term variations in the choice set to identify substitution patterns, even when the changes to the choice set are not fully observed.

We collect and analyze a dataset in which a new wireless technology allows for quite detailed information on sales and inventories. However, the method we describe can be used in any setting in which periodic information on sales is available with inventory or capacity data. For example, hospitals, airlines, and sporting or concert events often have fixed and/or observable capacities, and many retail markets collect periodic inventory data. When such data are available, researchers gain valuable information on short-run choice set variation. Our results in this paper indicate that accounting for that choice set variation can substantially reduce potential biases in standard estimates for some markets, and that researchers should take on the responsibility to adjust for the effects of product availability in demand estimation when possible.

Table 1: Summary of Products and Markups

Product	Category	% SO (Low/Up)		p	c	Share	ADS	Mach.
PopTart	Pastry	5.26	6.51	1.00	0.35	3.79	0.93	49
Choc Donuts	Pastry	18.43	21.45	1.00	0.46	3.28	0.79	53
Ding Dong	Pastry	13.71	15.97	1.00	0.46	3.14	0.72	54
Banana Nut Muffin	Pastry	7.07	8.28	1.00	0.40	2.94	0.65	52
Rice Krispies	Pastry	1.15	1.39	1.00	0.31	2.22	0.50	54
Pastry	Pastry	15.72	18.68	1.00	0.46	0.92	1.02	23
Gma Oatmeal Raisin	Cookie	1.19	1.48	0.75	0.23	3.10	0.68	52
Chips Ahoy	Cookie	0.67	0.89	0.75	0.25	2.80	0.60	53
Nutter Butter Bites	Cookie	0.19	0.25	0.75	0.26	1.97	0.45	51
Knotts Raspberry Cookie	Cookie	0.34	0.43	0.75	0.19	1.96	0.43	52
Gma Choc Chip	Cookie	1.45	1.84	0.75	0.22	1.50	0.80	45
Gma Mini Cookie	Cookie	5.01	5.59	0.75	0.21	0.52	0.54	46
Gma Caramel Choc Chip	Cookie	4.32	4.92	0.75	0.23	0.23	0.61	50
Rold Gold	Chips	6.77	8.15	0.90	0.27	4.56	0.96	54
Sunchip Harvest	Chips	5.35	6.45	0.90	0.27	4.36	0.92	54
Cheeto Crunchy	Chips	3.86	4.60	0.90	0.27	3.82	0.80	54
Dorito Nacho	Chips	1.74	2.13	0.90	0.27	3.71	0.80	54
Gardetto Snackens	Chips	0.93	1.08	0.75	0.23	2.71	1.04	54
Ruffles Cheddar	Chips	3.79	4.69	0.90	0.27	1.98	0.65	48
Fritos	Chips	2.23	2.53	0.90	0.27	1.93	0.44	51
Lays Potato Chip	Chips	2.43	2.78	0.90	0.27	1.83	0.40	54
Misc Chips 2	Chips	1.74	2.00	0.90	0.28	1.15	0.37	54
Dorito Guacamole	Chips	2.27	2.67	0.90	0.28	0.93	0.47	46
Munchies	Chips	5.12	5.59	0.90	0.25	0.60	0.50	42
Misc Chips 1	Chips	5.43	5.87	0.90	0.26	0.50	0.42	42
Munchies Hot	Chips	0.37	0.43	0.75	0.25	0.47	0.53	40
Snickers	Chocolate	0.61	0.82	0.75	0.33	9.21	1.92	54
Twix	Chocolate	0.52	0.67	0.75	0.33	6.88	1.43	54
M&M Peanut	Chocolate	1.38	1.75	0.75	0.33	5.22	1.14	52
Reese's Cup	Chocolate	0.61	0.72	0.75	0.33	2.70	0.57	54
Kit Kat	Chocolate	0.47	0.58	0.75	0.33	2.47	0.52	54
Caramel Crunch	Chocolate	0.41	0.50	0.75	0.33	2.44	0.51	54
Hershey Almond	Chocolate	0.28	0.33	0.75	0.33	1.71	0.40	50
M&M	Chocolate	3.59	3.98	0.75	0.33	0.55	0.50	48
Babyruth	Chocolate	3.56	3.97	0.75	0.28	0.22	0.29	52
Kar Nut Sweet/Salt	Candy	1.21	1.45	0.75	0.22	3.23	0.69	53
Snackwell	Candy	0.35	0.40	0.75	0.28	1.89	0.41	54
Skittles	Candy	0.72	0.95	0.75	0.34	1.56	0.79	47
Payday	Candy	0.35	0.44	0.75	0.35	1.27	0.53	54
Oreo	Candy	0.10	0.12	0.75	0.22	1.05	0.25	49
Peter Pan (Crck)	Candy	0.54	0.63	0.75	0.12	0.83	0.42	47
Peanuts	Candy	1.07	1.18	0.75	0.26	0.81	0.44	45
Starburst	Candy	2.85	3.62	0.75	0.33	0.79	0.80	47
Hot Tamales	Candy	4.58	5.33	0.75	0.27	0.27	0.46	54

Product and Category provided by the vending company, % SO (Low/Up) reports lower and upper bounds on stock-out frequencies, p is price charged at vending machines, c is wholesale cost, Share is 'inside good' marketshare, ADS is average daily sales across all machines that carried a product, and Mach. is number of machines that carried a product.

Table 2: Regression of Profit on Stock-Out Variables

	(1)	(2)	(3)
# Products Stocked Out	-0.895*** (0.017)	-0.437*** (0.021)	-0.408*** (0.023)
(max) SO's, category		-1.896*** (0.053)	
# SO, Pastry			-2.273*** (0.062)
# SO, Cookie			-2.637*** (0.21)
# SO, Chips			-1.450*** (0.069)
# SO, Chocolate			-1.611*** (0.15)
# SO, Candy			-3.847*** (0.24)
Constant	7.425*** (0.049)	7.727*** (0.050)	7.743*** (0.050)
Observations	111195	111195	111195
R^2	0.0238	0.0349	0.0367

An observation is a four-hour period at an individual machine (recorded wirelessly).

Table 3: Regressions of Stockout Rates on Starting Inventory

	OLS (1)	OLS (2)	OLS (3)	Probit (1)	Probit (2)	Probit (3)
Beginning Inv	-0.0168*** (0.00025)	-0.00872*** (0.00042)	-0.0101*** (0.00055)	-0.0142*** (0.00022)	-0.00681*** (0.00035)	-0.00893*** (0.00036)
Hours	0.00203*** (0.00011)	0.00217*** (0.00010)	0.00278*** (0.00011)	0.00169*** (0.000092)	0.00174*** (0.000081)	0.00212*** (0.000084)
Constant	0.339***	0.225***	0.238***			
FE	-	Product	Prod x Mach	-	Product	Prod + Mach
Observations	98900	98900	98900	98900	98900	98900
R^2	0.0486	0.1326	0.2092	0.0585	0.1561	0.1788

An observation is an individual product at a service visit for an individual machine.

Table 4: Non-linear Parameter Estimates

	Full Availability	Ignore	EM
Single Nesting Parameter (λ)			
Category	0.525 (0.029)	1.075 (0.020)	0.768 (0.0002)
Neg LL	2,279,524	1,906,749	2,268,233
Category-Specific Nesting Parameter (λ)			
Pastry	1.550 (0.101)	1.131 (0.026)	0.833 (0.021)
Cookie	0.518 (0.044)	0.677 (0.051)	0.520 (0.039)
Chips	0.202 (0.068)	1.465 (0.044)	0.805 (0.035)
Chocolate	0.673 (0.108)	0.767 (0.055)	0.465 (0.041)
Candy	0.332 (0.048)	0.492 (0.055)	0.475 (0.047)
Neg LL	2,279,496	1,906,631	2,268,096
Random Coefficients			
Fat	0.566 (1.409)	0.000 (0.048)	0.306 (0.793)
Salt	2.851 (0.045)	0.000 (0.047)	2.523 (0.010)
Sugar	5.638 (0.021)	0.000 (0.251)	4.822 (0.007)
Neg LL	2,279,544	1,906,756	2,267,203
Choice Sets	238	2649	3966
Marketsize	5,786,721	5,319,851	5,786,721

Full Availability assumes that all products stocked in a machine are available to all consumers (ie., it ignores stockout events). Ignore adjusts for stockouts during periods in which all sales and availability regimes are observed, but ignores (discards) periods in which stockouts happened at an unknown point in time. EM adjusts for all stockout events, regardless of whether the timing of a stockout was fully observed in the data. Standard errors are reported in parentheses. The nesting parameter λ is the parameter specified in McFadden (1978), rather than the parameter σ specified in Cardell (1997) or Berry (1994), where $\lambda \approx (1 - \sigma)$.

Table 5: Best Substitutes

Product	Category-Specific Nested Logit	Random Coefficients
PopTart	Choc Donuts	Snickers
Choc Donuts	PopTart	Snickers
Ding Dong	Choc Donuts	Snickers
Banana Nut Muffin	Choc Donuts	PopTart
Rice Krispies	Choc Donuts	Snickers
Gma Oatmeal Raisin	Gma Choc Chip	Banana Nut Muffin
Chips Ahoy	Gma Choc Chip	Snickers
Nutter Butter Bites	Gma Choc Chip	Snickers
Knotts Raspberry Cookies	Gma Choc Chip	Snickers
Gma Choc Chip	Gma Oatmeal Raisin	Snickers
Rold Gold	Sunchip Harvest	Cheeto Crunchy
Sunchip Harvest	Rold Gold	Rold Gold
Dorito Nacho	Rold Gold	Rold Gold
Cheeto Crunchy	Rold Gold	Rold Gold
Ruffles Cheddar	Rold Gold	Rold Gold
Fritos	Rold Gold	Rold Gold
Lays Potato Chip	Rold Gold	Rold Gold
Munchies Hot	Rold Gold	Rold Gold
Misc Chips 2	Rold Gold	Rold Gold
Munchies	Rold Gold	Rold Gold
Dorito Guacamole	Rold Gold	Rold Gold
Snickers	Twix	Twix
Twix	Snickers	Snickers
M&M Peanut	Snickers	Snickers
Reeses	Snickers	Snickers
Kit Kat	Snickers	Snickers
Caramel Crunch	Snickers	Snickers
M&M	Snickers	Snickers
Hershey Almond	Snickers	Snickers
Starburst	Skittles	Skittles
Kar Nut Sweet/Salt	Starburst	Snickers
Snackwell	Starburst	Snickers
Skittles	Starburst	Starburst
Oreo	Starburst	Snickers
Peanuts	Starburst	Rold Gold

Reports the closest substitute for each of the 35 most commonly-held products, based on EM-corrected demand estimates.

Table 6: Predicted Weekly Sales

	Category-Specific Nested Logit				Random Coefficients Logit			
	Full	Ignore	EM	Diff	Full	Ignore	EM	Diff
PopTart	9.08	8.92	9.49	0.41	8.75	8.8	9.21	0.46
Choc Donuts	8.75	9.85	10.63	1.88	8.50	9.77	10.47	1.96
Ding Dong	7.72	8.11	8.82	1.09	7.38	8.01	8.60	1.22
Banana Nut Muffin	6.67	6.57	7.04	0.37	6.54	6.47	6.91	0.37
Rice Krispies	4.99	4.57	4.90	-0.08	4.97	4.49	5.00	0.04
Gma Oatmeal Raisin	6.71	6.21	6.84	0.13	6.85	6.31	6.83	-0.02
Chips Ahoy	5.84	5.29	5.89	0.06	6.03	5.42	6.06	0.03
Nutter Butter Bites	4.28	3.88	4.27	-0.01	4.39	3.95	4.38	-0.01
Knott's Raspberry Cookie	4.18	3.83	4.18	0.00	4.26	3.90	4.24	-0.02
Gma Choc Chip	8.22	7.24	8.21	-0.01	7.75	6.93	7.68	-0.07
Rold Gold	9.89	9.60	10.69	0.80	10.06	9.55	10.87	0.81
Sunchip Harvest	9.43	8.92	9.97	0.54	9.37	8.86	9.99	0.62
Dorito Nacho	8.22	7.43	8.20	-0.02	8.17	7.36	8.24	0.07
Cheeto Crunchy	8.21	7.60	8.41	0.20	8.24	7.54	8.48	0.24
Ruffles Cheddar	6.91	6.25	6.89	-0.03	6.90	6.19	6.94	0.04
Fritos	4.56	4.06	4.55	-0.01	4.53	4.02	4.58	0.05
Lays Potato Chip	4.15	3.79	4.17	0.02	4.13	3.75	4.20	0.07
Munchies Hot	7.82	7.42	7.89	0.07	7.92	7.22	7.96	0.05
Misc Chips 2	3.42	3.16	3.50	0.08	3.43	3.11	3.53	0.10
Munchies	5.36	4.81	5.44	0.08	5.28	4.80	5.46	0.18
Dorito Guacamole	4.71	4.10	4.60	-0.11	4.61	4.10	4.61	0.01
Snickers	20.14	18.5	20.23	0.09	19.72	18.57	19.73	0.01
Twix	14.99	13.48	14.97	-0.02	14.70	13.54	14.68	-0.03
M&M Peanut	11.35	10.43	11.50	0.16	11.13	10.47	11.30	0.17
Reese's Cup	5.71	5.18	5.70	-0.02	5.67	5.20	5.66	-0.01
Kit Kat	5.23	4.72	5.21	-0.02	5.06	4.74	5.05	0.00
Caramel Crunch	5.19	4.74	5.16	-0.03	5.12	4.77	5.11	-0.01
M&M	5.76	5.28	5.82	0.07	5.65	5.26	5.66	0.01
Hershey Almond	3.99	3.64	3.97	-0.03	3.96	3.66	3.96	-0.01
Starburst	8.67	7.98	8.83	0.16	8.62	8.28	8.68	0.07
Kar Nut Sweet/Salt	6.45	5.95	6.60	0.15	6.84	6.28	6.88	0.04
Snackwell	3.81	3.53	3.92	0.11	4.07	3.74	4.12	0.04
Skittles	8.08	7.31	8.12	0.04	8.11	7.21	8.02	-0.09
Oreo	2.33	2.14	2.39	0.05	2.49	2.27	2.49	0.00
Peanuts	4.50	4.11	4.50	0.00	4.39	4.02	4.41	0.02
Total	245.34	228.59	251.53	6.19	243.59	228.56	250.00	6.41

Reports predicted weekly sales for each of the 35 most commonly-held products under the Full Availability, Ignore, and EM-corrected models. Diff reports the difference in weekly sales between the EM-corrected and Full Availability estimates for each of the two specifications (Category-specific Nested Logit and Random Coefficients Logit).

Table 7: Weekly Sales Impact of Simulated Stockout

	Category-Specific Nested Logit			Random Coefficients Logit		
	Full	Ignore	EM	Full	Ignore	EM
PopTart	-9.08	-8.92	-9.49	-8.75	-8.80	-9.21
Choc Donuts	-8.75	-9.85	-10.63	-8.50	-9.77	-10.47
Ding Dong	-2.13	-0.54	1.21	3.11	0.17	2.97
Banana Nut Muffin	-1.84	-0.44	0.97	4.02	0.14	3.60
Rice Krispies	-1.37	-0.31	0.67	0.56	0.10	0.49
Gma Oatmeal Raisin	-6.71	-6.21	-6.84	-6.85	-6.31	-6.83
Chips Ahoy	-5.84	-5.29	-5.89	-6.03	-5.42	-6.06
Nutter Butter Bites	1.41	0.88	1.42	0.56	0.09	0.48
Knott's Raspberry Cookie	1.38	0.87	1.40	0.94	0.08	0.77
Gma Choc Chip	2.71	1.64	2.74	1.83	0.15	1.49
Rold Gold	-9.89	-9.60	-10.69	-10.06	-9.55	-10.87
Sunchip Harvest	-9.43	-8.92	-9.97	-9.37	-8.86	-9.99
Dorito Nacho	2.46	-0.91	0.67	0.44	0.16	0.41
Cheeto Crunchy	2.45	-0.93	0.69	0.83	0.16	0.72
Ruffles Cheddar	2.07	-0.76	0.57	0.44	0.13	0.39
Fritos	1.36	-0.50	0.37	0.20	0.09	0.19
Lays Potato Chip	1.24	-0.46	0.34	0.19	0.08	0.18
Munchies Hot	2.34	-0.91	0.65	0.47	0.16	0.42
Misc Chips 2	1.02	-0.39	0.29	0.15	0.07	0.15
Munchies	1.60	-0.59	0.45	0.42	0.10	0.38
Dorito Guacamole	1.41	-0.50	0.38	0.16	0.09	0.15
Snickers	-20.14	-18.50	-20.23	-19.72	-18.57	-19.73
Twix	-14.99	-13.48	-14.97	-14.70	-13.54	-14.68
M&M Peanut	2.96	2.00	5.16	2.77	0.23	2.32
Reese's Cup	1.49	0.99	2.56	1.21	0.11	0.99
Kit Kat	1.37	0.90	2.34	1.67	0.10	1.38
Caramel Crunch	1.36	0.91	2.32	1.18	0.10	0.97
M&M	1.50	1.01	2.61	1.94	0.11	1.61
Hershey Almond	1.04	0.70	1.78	0.66	0.08	0.54
Starburst	-8.67	-7.98	-8.83	-8.62	-8.28	-8.68
Kar Nut Sweet/Salt	-6.45	-5.95	-6.60	-6.84	-6.28	-6.88
Snackwell	1.94	1.34	1.58	0.76	0.08	0.63
Skittles	4.10	2.78	3.27	4.51	0.16	3.91
Oreo	1.19	0.81	0.96	0.69	0.05	0.56
Peanuts	2.29	1.56	1.81	0.12	0.09	0.12

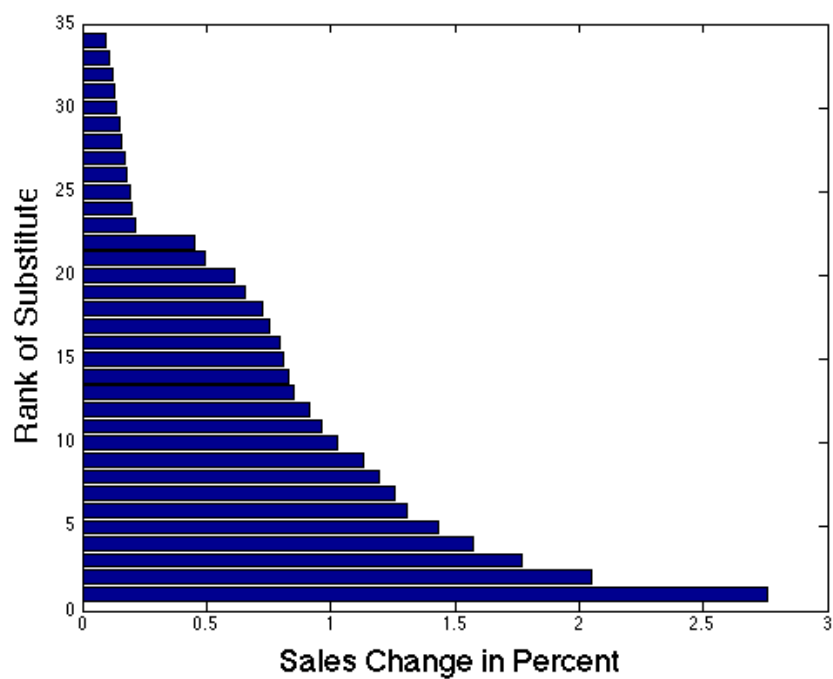
Reports predicted weekly sales impact on the 35 most commonly-held products of a simulated stockout of the top two selling products in each category (Chocolate Donuts, Strawberry Frosted PopTarts, Grandma's Oatmeal Raisin Cookie, Chips Ahoy Cookies, Rold Gold Pretzels, Sunchips Harvest Cheddar, Snickers, Twix, Starburst, and Kar Nut's Sweet & Salty Mix).

Table 8: Weekly Profit Impact of Simulated Stockout

	Category-Specific Nested Logit			Random Coefficients Logit		
	Full	Ignore	EM	Full	Ignore	EM
Foregone Sales of Stockout						
Pastry	-17.83	-18.77	-20.12	-17.25	-18.58	-19.68
Cookie	-12.55	-11.50	-12.74	-12.88	-11.73	-12.89
Chips	-19.32	-18.52	-20.66	-19.44	-18.41	-20.87
Chocolate	-35.13	-31.98	-35.20	-34.42	-32.11	-34.41
Candy	-15.13	-13.93	-15.44	-15.46	-14.56	-15.56
Total	-99.95	-94.69	-104.15	-99.45	-95.39	-103.4
Increased Sales of Substitutes						
Pastry	-5.34	-1.29	2.85	7.69	0.41	7.06
Cookie	5.51	3.38	5.56	3.33	0.32	2.73
Chips	15.95	-5.95	4.41	3.30	1.04	2.98
Chocolate	9.73	6.50	16.77	9.43	0.74	7.81
Candy	9.51	6.49	7.62	6.06	0.37	5.22
Total	35.35	9.14	37.20	29.81	2.88	25.81
Change in Total Sales						
Pastry	-23.17	-20.05	-17.27	-9.56	-18.16	-12.62
Cookie	-7.04	-8.12	-7.17	-9.55	-11.41	-10.16
Chips	-3.37	-24.46	-16.25	-16.14	-17.37	-17.88
Chocolate	-25.40	-25.48	-18.44	-25.00	-31.37	-26.6
Candy	-5.61	-7.44	-7.81	-9.39	-14.19	-10.34
Total	-64.6	-85.55	-66.95	-69.65	-92.50	-77.59
% Staying Inside						
Pastry	29.97	6.86	14.15	44.6	2.21	35.89
Cookie	43.88	29.42	43.68	25.83	2.73	21.20
Chips	82.55	32.10	21.33	16.95	5.65	14.29
Chocolate	27.69	20.33	47.63	27.38	2.30	22.7
Candy	62.88	46.60	49.38	39.23	2.56	33.55
Total	35.37	9.66	35.72	29.97	3.02	24.96
Change in Gross Profit						
Pastry	-13.72	-11.86	-10.27	-6.04	-10.76	-7.75
Cookie	-3.49	-4.08	-3.56	-4.80	-5.82	-5.12
Chips	-2.41	-15.30	-10.32	-10.22	-10.96	-11.32
Chocolate	-10.67	-10.70	-7.74	-10.50	-13.18	-11.17
Candy	-2.72	-3.54	-3.73	-4.62	-6.64	-5.04
Total	-33.01	-45.48	-35.62	-36.19	-47.36	-40.40

Reports predicted weekly profit impact of a simulated stockout of the top two selling products in each category (Chocolate Donuts, Strawberry Frosted PopTarts, Grandma’s Oatmeal Raisin Cookie, Chips Ahoy Cookies, Rold Gold Pretzels, SunChips Harvest Cheddar, Snickers, Twix, Starburst, and Kar Nut’s Sweet & Salty Mix). Uses the same group of 35 most commonly-held products in sales and sales impact tables.

Figure 1: Median Change in Sales of Substitutes by Rank



Reports the median change in sales of substitute by rank using the demand parameters from the EM-corrected random-coefficients specification.

A Appendix

A.1 Latent Types

As discussed in section 3.4, one might worry that our method relies too heavily on the assumption of conditional independence in some settings. For example, suppose we had daily data and we knew that there were very different consumer segments making purchases in the morning and the afternoon. If we knew a stockout happened sometime during the day, we might not expect the same distribution of consumer types before and after the stockout. This would generate latent consumer types, which are not addressed in our baseline model.

In this section, we discuss a method for using finite mixtures to incorporate latent types. We then show that when we add latent types into a model with latent stockout events, we require information on the joint distribution of the two sets of mixing parameters: those that apply to the latent types, and those that apply to the latent stockout events. We discuss the implication of this for alternative approaches that use simulation.

First consider how one might incorporate latent types in the absence of stock-out events. Suppose there are two consumer segments, type A and type B , with different tastes for a characteristic x . Each type has its own mean and standard deviation for its taste for x , denoted $\theta = [\mu_A, \mu_B, \sigma_A, \sigma_B]$. In addition, a mixing parameter, γ indicates the share of segment A in the population.³⁹ If we consider two products with different prices, and two consumer segments varying in their distaste for price, then the following choice probabilities apply for each type (assume non-subscripted elements are the full vector, and a random-coefficients logit specification).

$$\begin{aligned} p_{jtA}(a_t, x, d, \mu_A, \sigma_A) &= \int \frac{\exp(\beta_i x_{jt} + d_j)}{1 + \sum_{k \in a_t} \exp(\beta_i x_{kt} + d_k)} \phi(\beta_i | \mu_A, \sigma_A) \\ p_{jtB}(a_t, x, d, \mu_B, \sigma_B) &= \int \frac{\exp(\beta_i x_{jt} + d_j)}{1 + \sum_{k \in a_t} \exp(\beta_i x_{kt} + d_k)} \phi(\beta_i | \mu_B, \sigma_B) \end{aligned}$$

The resulting population shares would then be a mixture of the two:

$$p_{jt}(x, \theta) = \gamma \cdot p_{jtA}(a_t, x, d, \mu_A, \sigma_A) + (1 - \gamma) \cdot p_{jtB}(a_t, x, d, \mu_B, \sigma_B)$$

We could estimate this model using a variety of techniques, including ML, where $\theta = [\delta, \mu_A, \mu_B, \sigma_A, \sigma_B, \gamma]$. When p_{jt} is computed via simulation, then we use MSL.

³⁹In many datasets, one observes aggregate annual sales, and uses changes in average annual prices as the primary source of variation in choice sets. Standard models assume that all consumers face the same choice set in a particular year, and that they are exchangeable (or IID). It is easy to generate exceptions to this. For example, imagine an iPhone that costs \$600 for the first six months, and \$400 for the next six months. Annual sales are reported with an average annual price of \$500, and we cannot recover the relevant structural parameters for the early vs. late purchasers.

It may also seem appealing to include moments regarding observed substitution after stockouts and use this to improve our estimates. Recall the multinomial is a semi-parametric form; once we fix a parameterization for $p_{jt}(\theta)$, our focus is on pinning down the mapping, not fitting the cell counts. We are already using all of the stockout information in identification of the choice probabilities, and ML estimation achieves the semi-parametric efficiency bound.⁴⁰

Now we add latent stockouts to the estimation. Suppose we observe a machine at 4pm and at 8pm. There are still two types of customers: students and administrative staff, with different distributions of preferences for snack foods. Each β_i is drawn from either $\beta_i \sim (b_S, W_S)$ or $\beta_i \sim (b_A, W_A)$ for students and administrators respectively. We know that the staff all leave at 5pm sharp, and the students are around until at least 8pm. This scenario is a bit problematic, because we do not expect the same proportions of students and staff before and after the stockout, and they are thus non-exchangeable.

In cases where such phenomena are important, we could specify the random coefficients choice probabilities for students and staff before the stockout as $p_j(a_t, \theta_S)$, and $p_j(a_t, \theta_A)$ respectively, and after the stockout as $p_j(a_s, \theta_S)$, and $p_j(a_s, \theta_A)$. For a guess of $\theta = [\theta_A, \theta_S]$ we know all four choice probabilities. Now we have a four-element mixture with three mixing parameters; α governs before and after the stockout and γ 's indicate the share of each subpopulation before and after the stockout. Market shares are now given by:

$$p_{jt}(\theta, \alpha, \gamma) = \alpha [\gamma_t p_j(a_t, \theta_S) + (1 - \gamma_t) p_j(a_t, \theta_A)] + (1 - \alpha) [\gamma_s p_j(a_s, \theta_S) + (1 - \gamma_s) p_j(a_s, \theta_A)]$$

The same two approaches we used before are available. We can work with the “mixed” probability or we can impute the sales (sufficient statistics) for each of the four cases and use an EM-type procedure. In the baseline case without latent types, the model implies a distribution on α (through the conditional negative binomial distribution on sales), which is easy to integrate out. We could use a similar technique if we fixed:

$$\begin{aligned} \tilde{p}_j(a_t, \theta, \gamma) &= \gamma_t p_j(a_t, \theta_S) + (1 - \gamma_t) p_j(a_t, \theta_A) \\ \tilde{p}_j(a_s, \theta, \gamma) &= \gamma_s p_j(a_s, \theta_S) + (1 - \gamma_s) p_j(a_s, \theta_A) \end{aligned}$$

⁴⁰Relatedly, random-coefficients demand estimators like that in Berry, Levinsohn, and Pakes (1995) use a mapping between parameters and data to produce choice probabilities. Those estimators include an error in the space of latent utilities rather than in the space of choice probabilities (i.e., the product-specific unobservables, ξ_{jt} 's), and require that the MSM condition on choice probabilities holds exactly. This is essentially the same problem we face: under the “true” model, each consumer has her own β_i , but the data do not include information about this latent consumer type. Stockouts are another example where data relevant to the choice probability computation is unobserved, except now it is the sales under different availability regimes. Thus, the two types of estimators represent two ways to handle these sorts of problems. One possibility is to impute the missing sufficient statistics throughout the dataset (this is how we handle stockouts). The other is to specify a more complicated mixture form for choice probabilities, which integrates out the latent variable (this is how BLP-style estimators deal with random coefficients). Just as we can (and do) formulate the stockout problem as a mixture of choice probabilities across availability regimes, we could in fact write the random coefficients model as a (finite-mixture) missing data problem if we discretized the type-space. An example of a paper that does this (although not explicitly as such) is Bajari, Fox, and Ryan (2006).

While this procedure would work, we would lose smoothness in the choice probabilities, which makes optimization difficult.

Whether we choose to impute sales for all four cases, or work with the mixture likelihood, we need to make an assumption on the joint distribution of (α, γ) . While we knew the marginal distribution of α , we do not know this joint distribution without more data or assumptions. One could assume that all administrative staff leave at 5pm sharp, which would imply a joint distribution. Alternatively, one could add free parameters to the model. However, doing this several times for each stockout scenario would dramatically increase the parameter space. It should also be clear that simulating consumer utilities to generate probabilities is not going to solve this, because we still need to specify the joint distribution of (α, γ) .

A.2 Multiple Unobserved Stockouts

Addressing the case of multiple unobserved stockouts is quite similar to the single stockout case. The rest of the estimation procedure proceeds just as it did in the case of a single unobserved stockout, with the exception of the E-step (where the missing data are imputed). Conditional on the imputed values for the missing data, the M-step remains unchanged.

Suppose we denoted the single stockout case as a problem in which we integrated over the fraction of consumers facing a particular availability regime. Thus there are two availability regimes, which apply to fractions α and $(1 - \alpha)$ of the population, and one would write:

$$E[y_{jt}^s | \theta] = y_{jt} \int \frac{\alpha p_{jt}(\theta, a_s, x_t)}{\alpha p_{jt}(\theta, a_s, x_t) + (1 - \alpha) p_{jt}(\theta, a_t, x_t)} g(\alpha | \theta, \mathbf{y}_t) d\alpha$$

When moving to K unobserved stockouts, there are 2^K possible availability regimes, and $(2^K - 1)$ parameters $\boldsymbol{\alpha} = [\alpha_0, \alpha_A, \alpha_B, \dots, \alpha_{AB}, \dots]$ where subscripts denote stocked out products:

$$E[y_{jtl}^s | \theta] = y_{jt} \int \frac{\alpha_l p_{jt}(\theta, a_l, x_t)}{\sum_{\forall s} \alpha_s p_{jt}(\theta, a_s, x_t)} g(\boldsymbol{\alpha} | \theta, \mathbf{y}_t) d\boldsymbol{\alpha}$$

Once again we can evaluate the expectation exactly, by evaluating at every α in the domain, but this is now computationally much more difficult. If we think about the dimension of the problem, the order of the stockouts now matters (since different orders imply different choice probabilities). There are $K!$ ways to order K stockouts. Once we assume an ordering we must divide up M_t consumers among the K availability regimes. This means that the summation would require $\binom{M_t}{K}$ elements for each possible ordering of stockouts or $K! \times \binom{M_t}{K} = \frac{M_t!}{M_t - K!}$ elements overall. For M_t large and K small this is (roughly) approximated as M^K . If a week's worth of data contains $M_t = 1000$ and $K = 5$, this implies 10^{15} elements. In this case, approximate methods must be used to compute the expectation.

Suppose we try to compute $g(\boldsymbol{\alpha} | \cdot)$ for a single value of $\boldsymbol{\alpha}$. An important property to notice

is that each $\boldsymbol{\alpha}$ vector implies a unique sequence of stockouts. If we had $K = 2$ products then $\boldsymbol{\alpha} = [\alpha_0, \alpha_A, \alpha_B, \alpha_{AB}]$. In this case, we know that $[0.2, 0.3, 0, 0.5]$ implies that product A stocked out first, and then product B . Likewise $[0.5, 0, 0.1, 0.4]$ tells us that product B stocked out first and then product A . We might think about $\boldsymbol{\alpha} = [0.3, 0.2, 0.2, 0.3]$, but this is impossible since we could not have observed a sequence where A was available when B was not AND also observed B available when A was not. The probability of such a sequence is zero.

We define $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_K]$ as the beginning of period inventory for the K products that stocked out, where $(\boldsymbol{\omega}, \boldsymbol{\alpha})$ are arranged in the order of the stockouts. It is also helpful to write $a(\alpha^{(k)}) = a_k$ which denotes the availability set corresponding to the k^{th} component of $\boldsymbol{\alpha}$, and to define \mathbf{q} as the ordered sales vector for all stocked-out products, and $\mathbf{p}(\theta, a(\alpha), x_t)$ as the vector of choice probabilities with $(K + 1)$ elements where the first K elements are $p_k(\theta, a(\alpha), x_t)$ for the K products which stockout, and $p_{k+1} = \sum_{j \in a(\alpha)} p_j(\theta, a(\alpha), x_t)$, or the sum of the choice probabilities of all other goods (including the outside good). Finally we define an operator $h(\cdot)$ which takes a K dimensional vector and returns the vector of the last $K - 1$ elements.

Now we can define a finite recursive relationship for $g(\boldsymbol{\alpha}, \boldsymbol{\omega}, \theta, M_t)$, for the case where $K > 1$ in terms of the multinomial generalization of the negative binomial, the negative multinomial (defined below):

$$\begin{aligned} g(\boldsymbol{\alpha}, \boldsymbol{\omega}, \theta, M_t) &= \sum_{\forall h(\mathbf{q}): h(\mathbf{q} < \boldsymbol{\omega})} NMult\left(\alpha^{(1)} M_t, \omega^{(1)}, h(\mathbf{q}), \mathbf{p}(a(\alpha^{(1)}), \theta)\right) \cdot g(h(\boldsymbol{\alpha}), h(\boldsymbol{\omega} - \mathbf{q}), \theta, M_t) \\ &= \sum_{q^{(2)=0} \omega^{2-1}} \sum_{q^{(3)=0} \omega^{3-1}} \dots \sum_{q^{(K)=0} \omega^{K-1}} NMult\left(\alpha^{(1)} M_t, \omega^{(1)}, h(\mathbf{q}), \mathbf{p}(a(\alpha^{(1)}), \theta)\right) \cdot g(h(\boldsymbol{\alpha}), h(\boldsymbol{\omega} - \mathbf{q}), \theta, M_t) \end{aligned}$$

And for the base case $K = 1$ (all arguments scalar), it is identical to the single stockout case:

$$g(\alpha, \omega, M_t) = \frac{NegBin(\alpha M_t - \omega, \omega, p_j(a(\alpha), x_t, \theta))}{NegBinCDF(M_t, \omega, p_j(a(\alpha), x_t, \theta))}$$

In words, to evaluate the density at a vector of weights for different availability regimes, we “pop” the first element of the availability regime and inventory off of our stack, and compute the negative multinomial probability times the function $g(\cdot)$ applied to the remaining stack. Each time we call $g(\cdot)$ we must evaluate the sum over all possible sales configurations for the products that did not stock out during that particular availability regime, but eventually stocked out. This can still be computationally burdensome (and for some problems infeasible). This represents a dramatic savings because we need not worry about the ordering of products that did not stock out, thus there are only $K < J$ sums to evaluate.

Negative Multinomial

The negative multinomial is simply the multinomial generalization of the negative binomial. This entire family of distributions (binomial, multinomial, geometric, negative binomial, negative multinomial, etc.) are all just derived distributions for the Bernoulli process. We have results for multinomials, and geometrics, etc. because they frequently occur in applied problems, and these standard results are often incorporated in textbooks and statistical packages. The negative multinomial is a bit less common, and results are not as well known.

The negative multinomial is similar to the negative binomial in that it describes the probability of the number of failures r_t before ω_{kt} successes of the first cell are observed. What makes it different from the negative binomial is that it also accounts for q_2, \dots, q_K successes of the next $K - 1$ cells. The p.m.f. can be written:

$$NMult(r_t, \omega_{kt}, \mathbf{q}, \mathbf{p}(a, \theta)) = \frac{(r_t + \omega_{kt} + \sum_k q_{kt} - 1)!}{(r_t - \omega_{kt} - 1)! \omega_{kt}! \dots q_K! q_0!} p_1^{\omega_{kt}} p_2^{q_2} \dots p_k^{q_k} p_0^{r_t}$$

A.3 Alternative Computational Methods

For up to three unobservable stockouts, the exact method we present in the text is generally computationally feasible. However, when there are many stockouts at once, integrating over the distribution $g(\boldsymbol{\alpha}|\cdot)$ at all values of the domain becomes prohibitive, and approximate methods must be used. Recall the form of the E-step:

$$E[y_{jt}^s] = \sum_{\forall s} y_{jt} \frac{\alpha_s p_j(\theta, a_s, x_t)}{\sum_{\forall r} \alpha_r p_j(\theta, a_r, x_t)} g(\boldsymbol{\alpha}|\theta, y_{jt})$$

This can be easily approximated by linear functions since it is smooth and of the form $\frac{f(z_i)}{\sum_i f(z_i)} \in [0, 1]$, and because many stockouts do not induce large changes in $p_j(\cdot)$ (a stockout of Doritos often has very little effect on sales of Snickers). This means that quadrature, Monte Carlo integration, and Quasi-Monte Carlo integration should work fine even with a small number of points at which the function is evaluated. One way to generate random draws from $g(\cdot)$ is to pick an ordering for stockouts and then successively draw from the negative multinomial distribution, and repeat this for all possible orderings of the stockouts.

For extremely high dimensional problems we might find that even drawing from $g(\cdot)$ becomes too burdensome, as there are an increasing number of potential orderings for stockouts. One could consider simulating consumer purchases (because we know the sales are distributed as a multinomial for a given set of parameters θ). One only needs to track those products stocked out and an ‘other’ option. In this case, we simulate consumers until we’ve observed all of the stockouts. We can count the fraction of all consumers facing each availability set and compute α directly. One of the key benefits of our method is that it allows us to compute the expectation of the missing data by only evaluating over distributions of the stocked-out products, without resorting to computing the likelihood for every possible permutation of sales.

A.4 Identification Details for Random Coefficients Specification

Identification of nonlinear parameters in a random-coefficients specification comes from the fact that the sales of two products j, k will be differentially affected by a stockout of product l depending on how similar their characteristics x_j, x_k are to x_l . In the case where we have only one continuous, real-valued x_{jt} characteristic with a random coefficient, we can consider the choice probabilities for good j before and after the stockout of good l . That is availability regime $a'_t = \{a_t\} \setminus \{l\}$. The choice probabilities are:

$$\begin{aligned}
 p_{jt}(\theta) &= \frac{1}{ns} \sum_{i=1}^{ns} P_{ijt} = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp[d_j + \sigma v_i x_j]}{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]} \\
 \tilde{p}_{jt}(\theta) &= \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp[d_j + \sigma v_i x_j]}{1 + \sum_{k \in a'_t} \exp[d_k + \sigma v_i x_k]} \\
 &= \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp[d_j + \sigma v_i x_j]}{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]} \cdot \frac{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]}{1 + \sum_{k \in a'_t} \exp[d_k + \sigma v_i x_k]} \\
 &= \frac{1}{ns} \sum_{i=1}^{ns} \frac{P_{ijt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)}{1 - P_{ilt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)}
 \end{aligned}$$

Where the last equation follows because:

$$\begin{aligned}
 \frac{1 + \sum_{k \in a'_t} \exp[d_k + \sigma v_i x_k]}{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]} &= \frac{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]}{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]} - \frac{\exp[d_l + \sigma v_i x_l]}{1 + \sum_{k \in a_t} \exp[d_k + \sigma v_i x_k]} \\
 &= 1 - P_{ilt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)
 \end{aligned}$$

Specifically, for each consumer type i we inflate the probability of buying good j by a factor proportional to the probability that type i bought the stocked-out good l .⁴¹ Thus, we can think about a stockout as providing information not only about the level of p_{jt} , but also about the ratio of the choice probabilities before and after a stockout.

$$\frac{\tilde{p}_{jt}}{p_{jt}} = \frac{\sum_{i=1}^{ns} \frac{P_{ijt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)}{1 - P_{ilt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)}}{\sum_{i=1}^{ns} P_{ijt}(a_t, \mathbf{x}, \mathbf{d}, \sigma)}$$

Thus, identification comes from the fact that the correlation between choice probabilities (p_{jt}, p_{lt}) is determined by $\sigma(x_j - x_l)$.

⁴¹In the case of the plain logit model, P_{ilt} is constant across all types i , and we recover the IIA property so that all products are inflated by the same $\frac{1}{1-p_{lt}}$ factor.

A.5 Treating the Stockout Time as a Free Parameter

In section 3.2, we show that the likelihood for the dataset can be written down as the sum of the fully-observed and partially-observed observations, where the partially-observed observations must be integrated over the unobservable stockouts.

$$\hat{\theta} = \arg \max_{\theta} \sum_{\forall (a,x)} \left(\sum_{\forall t \in T_{mis}} E[y_{j,(a,x)}^{mis} | \theta] \ln p_j(\theta, a, x) + \sum_{\forall t \in T_{obs}} y_{j,(a,x)}^{obs} \ln p_j(\theta, a, x) \right)$$

This is the expected likelihood of the observed data $(\mathbf{a}, \mathbf{x}, \mathbf{y})$ given the parameter θ . Let α denote the fraction of consumers arriving before the stockout and $1 - \alpha$ denote the fraction of consumers arriving after the stockout. We could compute the expected sales for each product before and after the stockout:

$$\begin{aligned} E[y_{jt}^{before}] &= y_{jt} \int \frac{\alpha p_{jt}(\theta, a_t, x_t)}{\alpha p_{jt}(\theta, a_t, x_t) + (1 - \alpha) p_{jt}(\theta, a_s, x_t)} f(\alpha) \partial \alpha \quad \forall j \\ E[y_{jt}^{after}] &= y_{jt} - E[y_{jt}^{before}] \end{aligned}$$

This is essentially what we do in the E-Step of our EM procedure, though with a slightly different parameterization for the stockout time. An alternative might be to consider the marginal data augmentation framework of Tanner and Wong (1987), in which we think of the stockout time α as the missing data and estimate it as an additional parameter. In general this approach works when the integral is single dimensional because the integrand is a convex combination of choice probabilities. That is, there might exist a $\hat{\alpha}$ such that:

$$\frac{\hat{\alpha} p_{jt}(\theta, a_t, x_t)}{\hat{\alpha} p_{jt}(\theta, a_t, x_t) + (1 - \hat{\alpha}) p_{jt}(\theta, a_s, x_t)} = \int \frac{\alpha p_{jt}(\theta, a_t, x_t)}{\alpha p_{jt}(\theta, a_t, x_t) + (1 - \alpha) p_{jt}(\theta, a_s, x_t)} f(\alpha) \partial \alpha$$

If this were true we could treat $\hat{\alpha}$ as an additional parameter to estimate. Unfortunately, we don't have a single equation, but rather a set of $(J - 1)$ equations (one for each product that did not stock out), and only a single α . Thus only in very special (degenerate) cases can a single $\hat{\alpha}$ satisfy all $(J - 1)$ equations.⁴² This highlights the importance of always letting the E-Step operate on the sufficient statistics for estimation, rather than some other quantity. In our case, the sufficient statistics are sales under each regime, rather than stockout times. Our approach does follow the marginal data augmentation framework of Tanner and Wong (1987), but it works by considering a model where we know sales under all availability sets (even though these aren't directly observed), rather than integrating the likelihood at each guess of the parameters.

Standard approaches do not solve (5), but rather assume a different $f(\alpha)$ other than the

⁴²The independent poisson model as used by Anupindi, Dada, and Gupta (1998) is such a degenerate case.

one the stockout distribution implies. For example, in the case where we ignore the missing data it would be as if we set $f(\alpha) = 0$ everywhere that things are ambiguous (which is not a proper density distribution). Or in the case of full availability it would be as if we set $f(\alpha)$ to be a delta function that took on value 1 only at the full availability value of a_t . Assuming stockouts happen at the beginning or the end of the period places similar structure on $f(\alpha)$ (making it a delta function). The problem with this is that $f(\alpha)$ does not have any free parameters, but is completely specified by the demand parameters as a conditional negative binomial.

A.6 Standard Errors

The covariance matrix is just the inverse Fisher Information matrix or the negative expected Hessian, and for maximum likelihood estimators we use the outer product of scores in place of the Hessian, denoted as $H(\theta)$:

$$\begin{aligned}
 H(\theta) &= -E \left[\frac{\partial^2 L(\theta|Z)}{\partial \theta^2} \right] = -E \left[\frac{\partial l(\theta|Z)}{\partial \theta} \frac{\partial l(\theta|Z)'}{\partial \theta} \right] = \left[\sum_i \frac{\partial l_i(\theta|Z)}{\partial \theta} \times \frac{\partial l_i(\theta|Z)'}{\partial \theta} \right] \\
 l_i(\theta|y) &= y_{jt}(\theta) \ln p_{jt}(\theta)
 \end{aligned}$$

We obtain the score by differentiating the log-likelihood:

$$\begin{aligned}
 \frac{\partial l_i(\theta|y)}{\partial \theta} &= \frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta} \ln p_{jt}(\theta) + \underbrace{\hat{y}_{jt}(\theta) \frac{1}{p_{jt}(\theta)} \frac{\partial p_{jt}(\theta)}{\partial \theta}}_{\frac{\partial l_i(\theta|\hat{y}_{jt})}{\partial \theta}} \\
 &= \sum_{jt} \frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta} \ln p_{jt}(\theta) + \frac{\partial l_i(\theta|\hat{y}_{jt})}{\partial \theta}
 \end{aligned}$$

This gradient can be composed into two parts. The gradient of the observed data \hat{y}_{jt} is the sum of two components, the gradient where we assume we have the complete data, that is where \hat{y}_{jt} is treated as the truth, and a correction for the fact that the sufficient statistics are not fixed observable quantities but rather random quantities containing some uncertainty. We get the score as the gradient of the log-likelihood at the final parameter values, so all we need to do is compute the correction. The easiest way to actually compute this is to numerically differentiate the imputed sufficient statistics $\frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta}$

Note that were we to compute $\frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta}$, it would be:

$$\begin{aligned} \frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \left[\sum_{r_t \leq M_t} y_{jt} \frac{r_t \cdot p_{jt}(\theta)}{r_t \cdot p_{jt}(\theta) + (M_t - \omega_{kt} - r_t) \cdot p'_{jt}(\theta)} h(r_t | \omega_{kt}, p_k(\theta)) \right] \\ &= \frac{\widehat{y}_{jt}^0}{p_{jt}} \frac{\partial p_{jt}}{\partial \theta} + \widehat{y}_{jt}^0 b_t \frac{\partial p_{lt}}{\partial \theta} - \left(\frac{\partial p_{jt}}{\partial \theta} - \frac{\partial p'_{jt}}{\partial \theta} \right) \frac{\widehat{y}_{jt}^2}{p_{jt}} - (M_t - y_{jt}) \frac{\partial p'_{jt}}{\partial \theta} y_{jt} p_{jt} E_r \left[\frac{r_t}{D(r_t)^2} \right] \\ &+ \frac{\partial p_{lt}}{\partial \theta} y_{jt} p_{jt} E_r \left[\frac{r_t^2}{D(r_t)} \right] \end{aligned}$$

where

$$b_t = \left(\frac{2\omega_{kt}}{p_k} - \frac{M_t - \omega_{kt}}{1 - p_k} \right)$$

and

$$D(r_t) = r_t \cdot p_{jt}(\theta) + (M_t - \omega_{kt} - r_t) \cdot p'_{jt}(\theta)$$

This expression indicates that: (1) the quadratic term dies out as gradient of the choice probabilities becomes more similar before and after the stockout (which we could think about as the finite difference of the stockout), and (2) the other terms depend on how quickly the choice probabilities change in the parameters for the stocked-out product and the product of interest, roughly weighted by how much data is missing y_{jt} .⁴³

Delta Method Correction for Nested Logit Parameters

For the nested logit models, we require a correction when we report the standard errors for the product dummies, the d_j 's. We make the substitution $\tilde{\delta}_j = \frac{d_j}{\lambda_{j,k}}$ in estimation, where we define $\lambda_{j,k}$ as the nesting parameter which corresponds to the j th product. We write $\theta = [\tilde{\delta}_1, \dots, \tilde{\delta}_J, \lambda_1, \dots, \lambda_K] \in \mathbb{R}^L$, and use the delta method to recover standard errors on the d_j 's. The asymptotic distribution for θ is:

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, V(\hat{\theta}))$$

⁴³For those interested, the denominator term $D(v_t)$ is an incomplete beta function and Stirling's formula can be used to approximate derivatives for large M and small ω_{kt} .

The asymptotic distribution of our normalized d_j 's is the asymptotic distribution of $g(\theta)$, which we define as:

$$g(\theta) = \left[\tilde{\delta}_1 \lambda_{1k}, \dots, \tilde{\delta}_J \lambda_{Jk}, \lambda_1, \dots, \lambda_K \right]'$$

$$\sqrt{n} \left(g(\hat{\theta}) - g(\theta_0) \right) \sim N \left(0, \left[\frac{dg(\theta)}{d\theta} \right]' V(\hat{\theta}) \left[\frac{dg(\theta)}{d\theta} \right] \right)$$

To recover the marginal distributions of $g(\theta)$ we consider the function $g_j(\theta) = \tilde{\delta}_j \lambda_{jk}$, which has the following derivative vector $\nabla g_j = [0, \dots, 0, \lambda_{jk}, 0, \dots, 0, \tilde{\delta}_j, 0]$ where the nonzero elements are in position j and the position of the corresponding λ_k .

When we expand out we get the following quadratic form:

$$\sqrt{n} \left(g_j(\hat{\theta}) - g_j(\theta_0) \right) \sim N \left(0, \lambda_{jk}^2 \sigma_{\delta_j}^2 + \delta_j^2 \sigma_{\lambda_{jk}}^2 + 2\lambda_{jk} \delta_j \sigma_{\delta_j, \lambda_{jk}} \right)$$

And the standard error is:

$$SE(d_j) = \sqrt{\lambda_{jk}^2 \sigma_{\delta_j}^2 + \delta_j^2 \sigma_{\lambda_{jk}}^2 + 2\lambda_{jk} \delta_j \sigma_{\delta_j, \lambda_{jk}}}$$

For the λ parameters, no correction is necessary.

A.7 Additional Results

Tables 9, 10 and 11 report the estimates of the product dummies from each of the models estimated on the base dataset. Table 12 reports results of a second-stage regression of the fitted coefficients on product dummies on observable product characteristics to provide the mean levels of the tastes for these characteristics. The R^2 from these regressions is relatively low in the case of the nested-logit models. This indicates that the size of the unobservable ξ_j is large in our application, and highlights the need for product dummies. The R^2 in the single λ case is about 0.25, but this doubles to about 0.50 when we use category-specific λ 's. In a second-stage regression that includes category dummies, the R^2 improves significantly. The random-coefficients model allows for greater variation in the fitted product dummies, and has a higher R^2 in the second-stage regression.

Table 9: d_j Parameters, Single-Parameter Nested Logit

	Full	(S.E.)	Ignore	(S.E.)	EM	(S.E.)
PopTart	-5.488	0.040	-6.286	0.029	-5.754	0.022
Choc Donuts	-5.519	0.041	-6.177	0.028	-5.672	0.021
Ding Dong	-5.582	0.045	-6.388	0.031	-5.811	0.023
Banana Nut Muffin	-5.652	0.049	-6.616	0.035	-5.981	0.026
Rice Krispies	-5.801	0.057	-7.007	0.042	-6.255	0.032
Pastry	-5.576	0.047	-6.452	0.038	-5.860	0.028
Gma Oatmeal Raisin	-5.751	0.041	-6.623	0.030	-6.058	0.023
Chips Ahoy	-5.824	0.045	-6.784	0.033	-6.167	0.025
Nutter Butter Bites	-5.987	0.054	-7.127	0.040	-6.413	0.030
Knotts Raspberry Cookie	-5.999	0.054	-7.140	0.040	-6.429	0.030
Gma Choc Chip	-5.646	0.039	-6.535	0.034	-5.959	0.024
Gma Mini Cookie	-5.834	0.044	-6.700	0.035	-6.151	0.026
Gma Caramel Choc Chip	-5.804	0.043	-6.539	0.037	-6.050	0.028
Rold Gold	-5.120	0.057	-6.248	0.039	-5.510	0.030
Sunchip Harvest	-5.144	0.059	-6.328	0.041	-5.563	0.031
Dorito Nacho	-5.217	0.063	-6.527	0.044	-5.711	0.034
Cheeto Crunchy	-5.217	0.063	-6.501	0.044	-5.691	0.033
Gardetto Snackens	-4.961	0.049	-6.026	0.036	-5.348	0.027
Ruffles Cheddar	-5.307	0.068	-6.714	0.048	-5.842	0.037
Fritos	-5.525	0.080	-7.178	0.057	-6.155	0.043
Lays Potato Chip	-5.574	0.082	-7.252	0.058	-6.221	0.044
Munchies Hot	-5.238	0.064	-6.545	0.046	-5.742	0.035
Misc Chips 2	-5.674	0.088	-7.452	0.063	-6.353	0.047
Munchies	-5.443	0.076	-6.989	0.055	-6.020	0.041
Misc Chips 1	-5.438	0.075	-6.925	0.053	-5.992	0.040
Dorito Guacamole	-5.511	0.079	-7.159	0.059	-6.146	0.044
Snickers	-4.746	0.036	-5.533	0.026	-5.029	0.020
Twix	-4.900	0.045	-5.872	0.032	-5.255	0.024
M&M Peanut	-5.046	0.053	-6.148	0.037	-5.454	0.028
Reese's Cup	-5.406	0.072	-6.901	0.052	-5.983	0.039
Kit Kat	-5.452	0.075	-7.001	0.054	-6.051	0.041
Caramel Crunch	-5.456	0.075	-6.994	0.053	-6.057	0.041
M&M	-5.400	0.072	-6.893	0.053	-5.975	0.040
Hershey Almond	-5.593	0.083	-7.280	0.059	-6.256	0.045
Babyruth	-5.625	0.084	-7.272	0.062	-6.262	0.047
Starburst	-5.531	0.037	-6.344	0.028	-5.817	0.021
Kar Nut Sweet/Salt	-5.683	0.045	-6.638	0.033	-6.031	0.025
Snackwell	-5.959	0.060	-7.195	0.044	-6.422	0.033
Skittles	-5.578	0.041	-6.499	0.034	-5.904	0.025
Payday	-5.835	0.052	-6.892	0.038	-6.239	0.029
Oreo	-6.215	0.074	-7.733	0.055	-6.795	0.041
Peanuts	-5.885	0.058	-7.129	0.048	-6.352	0.035
Peter Pan (Crck)	-5.986	0.060	-7.201	0.045	-6.451	0.034
Hot Tamales	-5.809	0.052	-6.674	0.042	-6.100	0.032

Full assumes that all products stocked in a machine are available to all consumers (ie., it ignores stockout events). Ignore adjusts for stockouts during periods in which all sales and availability regimes are observed, but ignores (discards) periods in which stockouts happened at an unknown point in time. EM adjusts for all stockout events, regardless of whether the timing of a stockout was fully observed in the data. Standard errors are reported in parentheses.

Table 10: d_j Parameters, Category-Specific Nested Logit

	Full	(S.E.)	Ignore	(S.E.)	EM	(S.E.)
PopTart	-6.880	0.142	-6.361	0.037	-5.861	0.029
Choc Donuts	-6.937	0.144	-6.250	0.037	-5.767	0.028
Ding Dong	-7.126	0.156	-6.469	0.041	-5.923	0.031
Banana Nut Muffin	-7.348	0.173	-6.708	0.045	-6.110	0.035
Rice Krispies	-7.789	0.203	-7.118	0.055	-6.411	0.042
Pastry	-7.192	0.167	-6.539	0.047	-5.982	0.036
Gma Oatmeal Raisin	-5.743	0.063	-6.066	0.071	-5.732	0.054
Chips Ahoy	-5.816	0.069	-6.174	0.079	-5.810	0.059
Nutter Butter Bites	-5.976	0.083	-6.385	0.095	-5.978	0.072
Knotts Raspberry Cookie	-5.989	0.084	-6.393	0.096	-5.988	0.073
Gma Choc Chip	-5.638	0.060	-5.962	0.071	-5.637	0.052
Gma Mini Cookie	-5.826	0.067	-6.166	0.073	-5.820	0.057
Gma Caramel Choc Chip	-5.796	0.065	-6.057	0.067	-5.748	0.053
Rold Gold	-4.474	0.137	-7.004	0.087	-5.609	0.067
Sunchip Harvest	-4.483	0.140	-7.112	0.090	-5.665	0.069
Dorito Nacho	-4.511	0.150	-7.378	0.097	-5.822	0.076
Cheeto Crunchy	-4.511	0.150	-7.346	0.097	-5.801	0.075
Gardetto Snackens	-4.410	0.116	-6.714	0.080	-5.437	0.060
Ruffles Cheddar	-4.546	0.161	-7.632	0.105	-5.963	0.082
Fritos	-4.630	0.190	-8.265	0.125	-6.296	0.096
Lays Potato Chip	-4.649	0.196	-8.364	0.127	-6.367	0.099
Munchies Hot	-4.521	0.153	-7.381	0.097	-5.854	0.077
Misc Chips 2	-4.688	0.209	-8.634	0.136	-6.507	0.105
Munchies	-4.598	0.179	-8.016	0.119	-6.153	0.091
Misc Chips 1	-4.597	0.178	-7.901	0.113	-6.121	0.089
Dorito Guacamole	-4.624	0.188	-8.252	0.127	-6.287	0.097
Snickers	-4.935	0.138	-5.144	0.069	-4.664	0.052
Twix	-5.134	0.170	-5.387	0.087	-4.804	0.064
M&M Peanut	-5.322	0.200	-5.583	0.101	-4.927	0.075
Reese's Cup	-5.784	0.274	-6.120	0.139	-5.253	0.104
Kit Kat	-5.844	0.284	-6.191	0.144	-5.295	0.107
Caramel Crunch	-5.848	0.284	-6.187	0.144	-5.299	0.108
M&M	-5.779	0.276	-6.104	0.140	-5.243	0.103
Hershey Almond	-6.025	0.313	-6.390	0.159	-5.422	0.118
Babyruth	-6.059	0.314	-6.410	0.157	-5.437	0.119
Starburst	-5.286	0.061	-5.595	0.070	-5.440	0.060
Kar Nut Sweet/Salt	-5.384	0.075	-5.740	0.085	-5.574	0.074
Snackwell	-5.559	0.100	-5.997	0.113	-5.813	0.098
Skittles	-5.309	0.066	-5.638	0.077	-5.479	0.067
Payday	-5.486	0.088	-5.893	0.099	-5.718	0.086
Oreo	-5.721	0.123	-6.245	0.141	-6.041	0.122
Peanuts	-5.503	0.094	-5.923	0.109	-5.749	0.095
Peter Pan (Crck)	-5.583	0.102	-6.037	0.115	-5.848	0.100
Hot Tamales	-5.470	0.087	-5.814	0.090	-5.637	0.078

Full assumes that all products stocked in a machine are available to all consumers (ie., it ignores stockout events). Ignore adjusts for stockouts during periods in which all sales and availability regimes are observed, but ignores (discards) periods in which stockouts happened at an unknown point in time. EM adjusts for all stockout events, regardless of whether the timing of a stockout was fully observed in the data. Standard errors are reported in parentheses.

Table 11: d_j Parameters-Random Coefficients

	Full	(S.E.)	Ignore	(S.E.)	EM	(S.E.)
PopTart	-12.697	0.038	-6.185	0.010	-11.354	0.012
Choc Donuts	-8.946	0.033	-6.080	0.011	-8.111	0.006
Ding Dong	-11.116	0.030	-6.279	0.012	-9.965	0.007
Banana Nut Muffin	-14.200	0.072	-6.492	0.012	-12.615	0.019
Rice Krispies	-7.760	0.002	-6.858	0.014	-7.507	0.001
Pastry	-10.923	0.045	-6.335	0.021	-9.850	0.009
Gma Oatmeal Raisin	-11.614	0.023	-6.518	0.012	-10.564	0.008
Chips Ahoy	-8.374	0.007	-6.669	0.013	-7.940	0.002
Nutter Butter Bites	-8.072	0.002	-6.987	0.015	-7.785	0.001
Knotts Raspberry Cookie	-9.213	0.018	-6.999	0.015	-8.678	0.004
Gma Choc Chip	-8.875	0.010	-6.424	0.016	-8.299	0.002
Gma Mini Cookie	-7.552	0.003	-6.603	0.022	-7.276	0.001
Gma Caramel Choc Chip	-10.735	0.022	-6.450	0.027	-9.746	0.006
Rold Gold	-9.470	0.177	-6.103	0.010	-8.728	0.033
Sunchip Harvest	-6.672	0.005	-6.178	0.010	-6.480	0.001
Dorito Nacho	-6.964	0.009	-6.363	0.011	-6.790	0.003
Cheeto Crunchy	-8.038	0.034	-6.339	0.011	-7.628	0.010
Gardetto Snackens	-7.804	0.053	-5.895	0.011	-7.385	0.011
Ruffles Cheddar	-7.589	0.026	-6.537	0.012	-7.317	0.007
Fritos	-7.502	0.029	-6.969	0.015	-7.327	0.007
Lays Potato Chip	-7.667	0.026	-7.038	0.015	-7.471	0.006
Munchies Hot	-7.147	0.010	-6.383	0.017	-6.949	0.003
Misc Chips 2	-7.622	0.005	-7.225	0.018	-7.475	0.002
Munchies	-8.026	0.020	-6.792	0.018	-7.706	0.005
Misc Chips 1	-7.369	0.006	-6.736	0.019	-7.147	0.002
Dorito Guacamole	-7.196	0.005	-6.950	0.021	-7.109	0.001
Snickers	-9.014	0.018	-5.438	0.007	-8.225	0.005
Twix	-9.078	0.017	-5.754	0.008	-8.329	0.004
M&M Peanut	-8.914	0.015	-6.011	0.009	-8.229	0.004
Reese's Cup	-8.848	0.010	-6.712	0.013	-8.332	0.002
Kit Kat	-10.826	0.025	-6.804	0.013	-9.942	0.006
Caramel Crunch	-9.267	0.010	-6.798	0.013	-8.690	0.003
M&M	-10.947	0.026	-6.701	0.015	-10.022	0.007
Hershey Almond	-8.782	0.011	-7.064	0.015	-8.348	0.002
Babyruth	-11.926	0.031	-7.065	0.029	-10.877	0.009
Starburst	-12.602	0.050	-6.247	0.011	-11.301	0.015
Kar Nut Sweet/Salt	-8.424	0.009	-6.522	0.012	-7.949	0.002
Snackwell	-8.824	0.004	-7.040	0.015	-8.383	0.002
Skittles	-15.042	0.083	-6.384	0.015	-13.346	0.026
Payday	-8.970	0.009	-6.768	0.018	-8.452	0.003
Oreo	-10.383	0.007	-7.541	0.020	-9.736	0.003
Peanuts	-6.985	0.014	-6.969	0.021	-6.939	0.003
Peter Pan (Crck)	-8.099	0.019	-7.056	0.022	-7.849	0.004
Hot Tamales	-12.078	0.038	-6.570	0.031	-10.818	0.011

Full assumes that all products stocked in a machine are available to all consumers (ie., it ignores stockout events). Ignore adjusts for stockouts during periods in which all sales and availability regimes are observed, but ignores (discards) periods in which stockouts happened at an unknown point in time. EM adjusts for all stockout events, regardless of whether the timing of a stockout was fully observed in the data. Standard errors are reported in parentheses.

Table 12: d_j 's on Characteristics

	Single Parameter Nested			Category-Specific Nested			Random Coefficients		
	Full	Ignore	EM	Full	Ignore	EM	Full	Ignore	EM
Constant	-5.62 (0.19)	-7.25 (0.27)	-6.27 (0.21)	-4.61 (0.41)	-6.81 (0.37)	-5.72 (0.20)	-6.30 (0.39)	-7.04 (0.25)	-6.41 (0.35)
Calories	5.09 (2.21)	6.55 (3.05)	5.73 (2.42)	7.13 (4.76)	-3.03 (4.26)	2.15 (2.27)	7.85 (4.53)	6.35 (2.88)	7.71 (3.97)
Fat	-2.59 (1.09)	-3.37 (1.50)	-2.91 (1.19)	-4.82 (2.34)	0.37 (2.09)	-1.65 (1.12)	-3.01 (2.29)	-3.26 (1.42)	-3.06 (1.95)
Sodium	0.09 (0.41)	0.07 (0.56)	0.10 (0.45)	-0.66 (0.88)	0.38 (0.79)	0.14 (0.42)	-3.73 (0.84)	0.08 (0.53)	-3.01 (0.73)
Carbs	-2.47 (1.23)	-2.33 (1.70)	-2.40 (1.35)	-2.48 (2.66)	0.53 (2.38)	-1.65 (1.27)	-1.91 (2.53)	-2.35 (1.61)	-2.00 (2.21)
Sugar	-0.14 (0.50)	-0.01 (0.69)	-0.09 (0.55)	-2.30 (1.08)	2.29 (0.96)	0.73 (0.51)	-8.88 (1.02)	-0.02 (0.65)	-7.17 (0.90)
Chocolate	0.17 (0.14)	0.21 (0.20)	0.19 (0.15)	0.27 (0.30)	0.45 (0.27)	0.39 (0.15)	0.84 (0.29)	0.20 (0.18)	0.75 (0.25)
Cheese	0.18 (0.16)	0.17 (0.22)	0.18 (0.18)	0.51 (0.35)	0.02 (0.31)	0.17 (0.17)	-0.19 (0.33)	0.17 (0.21)	-0.12 (0.29)
R^2	0.218	0.253	0.219	0.525	0.608	0.480	0.917	0.244	0.900

Regresses the estimated linear parameters (d_j 's) on observed product characteristics for the Full, Ignore, and EM-corrected models. Standard errors are reported in parentheses.

References

- ACKERBERG, D., AND M. RYSMAN (2005): “Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects,” *RAND Journal of Economics*, 36(4).
- AGUIRREGABIRIA, V. (1999): “The Dynamics of Markups and Inventories in Retailing Firms,” *Review of Economic Studies*, 66, 278–308.
- ALLENBY, G., Y. CHEN, AND S. YANG (2003): “Bayesian Analysis of Simultaneous Demand and Supply,” *Quantitative Marketing and Economics*, 1, 251–304.
- ANUPINDI, R., M. DADA, AND S. GUPTA (1998): “Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products,” *Marketing Science*, 17(4), 406–423.
- ATHEY, S., AND G. IMBENS (2007): “Discrete Choice Models with Multiple Unobserved Choice Characteristics,” *International Economic Review*, 48(4).
- BAJARI, P., J. FOX, AND S. RYAN (2006): “Evaluating Wireless Consolidation Using Semiparametric Demand Estimation,” Working Paper.
- BALACHANDER, S., AND P. FARQUHAR (1994): “Gaining More by Stocking Less: A Competitive Analysis of Product Availability,” *Marketing Science*, 13(1), 3–22.
- BERRY, S. (1992): “An Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(60), 889–917.
- (1994): “Estimating discrete-choice models of product differentiation,” *RAND Journal of Economics*, 25(2), 242–261.
- BERRY, S., AND P. HAILE (2008): “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” Working Paper.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.
- (2004): “Automobile Prices in Market Equilibrium,” *Journal of Political Economy*, 112(1), 68–105.
- BRUNO, H. A., AND N. VILCASSIM (2008): “Structural Demand Estimation with Varying Product Availability,” *Marketing Science*, 27(6), 1126–1131.
- CARDELL, N. S. (1997): “Variance Component Structures for the Extreme Value and Logistic Distributions,” *Econometric Theory*, 13(2), 185–213.
- CARLTON, D. (1978): “Market Behavior with Demand Uncertainty and Price Inflexibility,” *American Economic Review*, 68, 571–587.

- CHINTAGUNTA, P., J. DUBE, AND K.-Y. GOH (2005): “Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models,” *Management Science*, 51(5), 832–849.
- DANA, J. (2001): “Competition in Price and Availability when Availability is Unobservable,” *Rand Journal of Economics*, 32(3), 497–513.
- DEMPSTER, A. P., N. M. LAIRD, AND D. B. RUBIN (1977): “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of the Royal Statistical Society*, 39(1), 1–38.
- DENECKERE, R., AND J. PECK (1995): “Competition Over Price and Service Rate When Demand is Stochastic: A Strategic Analysis,” *The RAND Journal of Economics*, 26(1), 148–162.
- DRAGANSKA, M., AND D. JAIN (2004): “A Likelihood Approach to Estimating Market Equilibrium Models,” *Management Science*, 50(5), 605–616.
- FOX, J. (2007): “Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices,” Working Paper.
- FOX, J., AND A. GANDHI (2008): “Identifying Heterogeneity in Economic Choice and Selection Models Using Mixtures,” Working Paper.
- GOLDBERG, P. K. (1995): “Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry,” *Econometrica*, 63(4), 891–951.
- GUPTA, S., P. CHINTAGUNTA, A. KAUL, AND D. R. WITTINK (1995): “Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data,” *Journal of Marketing Research*, 33(4), 383–398.
- HADLEY, G., AND T. WHITMAN (1963): *Analysis of Inventory Systems*. Prentice Hall.
- HARTLEY, H. (1958): “Maximum Likelihood from Incomplete Data,” *Biometrics*, 14(2), 174–194.
- LESLIE, P. (2004): “Price Discrimination in Broadway Theater,” *Rand Journal of Economics*, 35(3), 520–541.
- MATZKIN, R. (1992): “Nonparametric and Distribution-Free Estimation of Binary Threshold Crossing Models,” *Econometrica*, 60(2), 239–270.
- MCCARTHY, J., AND E. ZAKRAJSEK (forthcoming): “Inventory Dynamics and business cycles: What Has Changed?,” *Journal of Money, Credit, and Banking*.
- McFADDEN, D. (1978): “Modelling the Choice of Residential Location,” in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. North-Holland.

- McFADDEN, D., AND K. TRAIN (2000): “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics*, 15, 447–470.
- MORTIMER, J. H. (2008): “Vertical Contracts in the Video Rental Industry,” *Review of Economic Studies*, 75, 165–199.
- MUSALEM, A., M. OLIVARES, E. BRADLOW, C. TERWIESCH, AND D. CORSTEN (2008): “Structural Estimation of the Effect of Out-of-Stocks,” Working Paper.
- NARAYANAN, V., AND A. RAMAN (2004): “Aligning Incentives in Supply Chains,” *Harvard Business Review*, 82(11).
- NEVO, A. (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, 69, 307–342.
- NEVO, A., AND I. HENDEL (2007): “Measuring Implications of Sales and Consumer Inventory Behavior,” *Econometrica*.
- PETRIN, A. (2002): “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, 110(4), 705–729.
- TANNER, M., AND W. H. WONG (1987): “The Calculation of Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, 82(398).
- WWW.VENDING.ORG (2008): “Vending 101,” www.vending.org/industry/vending101.pdf.