NBER WORKING PAPER SERIES

DEMAND ESTIMATION UNDER INCOMPLETE PRODUCT AVAILABILITY

Christopher T. Conlon
Julie Holland Mortimer

Working Paper 14315
http://www.nber.org/papers/w14315

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2008

We thank Dan Ackerberg, Susan Athey, Steve Berry, Uli Doraszelski, J.P. Dube, Phil Haile, Wes Hartmann, Ken Hendricks, Guido Imbens, Phillip Leslie, Richard Mortimer, and Ariel Pakes 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), Harvard University, London School of Economics, the Stern School of Business at New York University, and Stanford University for helpful comments. Financial support for this research was generously provided through NSF grant SES-0617896. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Demand Estimation Under Incomplete Product Availability
Christopher T. Conlon and Julie Holland Mortimer
NBER Working Paper No. 14315
September 2008
JEL No. L0

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 dataset from a wireless inventory management systems, 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 in demand, and as a result, the corrected model predicts significantly larger impacts of stock-outs on profitability.

Christopher T. Conlon
Yale University
Department of Economics
New Haven, CT 06511
christopher.conlon@yale.edu

Julie Holland Mortimer
Harvard University
Department of Economics
Cambridge, MA 02138
and NBER
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 12 - 16 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 choices, restocking decisions or inventory policies in markets where stock-outs matter re-

[1]Relatedly, 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.

lies 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]

As a result of the enhanced data collection abilities that the wireless vending network provides, we observe a new source of variation in consumer choice sets. Namely, we observe actual stock-out events, which randomly change the set of products available at some locations for a period of time. This variation provides a new and 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.

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-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,

[2]Such 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).

[3]McCarthy 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.

[4]Note 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.

Laird, and Rubin 1977) to estimate the allocation of sales across the unobserved choice set regimes.

Relationship to Literature

The differentiated products literature in Industrial Organization (IO) has been primarily focused on two methodological problems. The first is the endogeneity of prices (Berry 1994), and the second is the determination of accurate substitution patterns. Berry, Levinsohn, and Pakes (1995) use unobserved product quality and unobserved tastes for product characteristics to more flexibly and accurately predict substitution patterns. The fundamental source of identification in these models comes through variation in choice sets across markets, typically through the price. Nevo (2001) uses a similar model to study a retail environment in his analysis of the market for Ready to Eat (RTE) Cereal. Further work (Petrin 2002, Berry, Levinsohn, and Pakes 2004) has focused on using interactions of consumer observables and product characteristics to better estimate substitution patterns. Berry, Levinsohn, and Pakes (2004) extend this idea even further and use second choice data from surveys in which consumers are asked which product they would have purchased if their original choice was unavailable. This paper's approach is a bit different because consumer-level stated second-choice data are unobserved, and substitution patterns are instead inferred from revealed substitution by exploiting short-run variations in the set of available choices. Recently, there have been several attempts made to present a fully Bayesian model of discrete choice consumer demand, among them Athey and Imbens (2007). While our paper uses a common Bayesian technique to address missing data, it is not a fully Bayesian model.

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 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. Estimating too many parameters is avoided by assuming that consumers leave the machine if their first two choices are unavailable. ADG did not observe the stock-out time and used EM techniques to estimate the Poisson model in the presence of missing choice-set data. 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. This paper aims to connect these literatures, by using modern differentiated product estimation techniques to obtain accurate estimates of substitution patterns while reducing the parameter space and applying missing-data techniques to correct these estimates for stockout-based substitution.

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. As these data become 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.

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. Often, such models start by deriving consumer choice probabilities from a random utility maximization (RUM) framework in a discrete choice setting (i.e., each consumer purchases exactly one unit of one product). Typically, the RUM problem is specified so that consumer choice probabilities take a logit form, and these choice probabilities are then used as inputs into a multinomial distribution that is estimated via GMM or maximum likelihood. Here, we start with a multinomial distribution for any general set of consumer choice probabilities (denoted p_{ij} for consumer i and product j) and derive a model of consumer sales. We use this focus in order to present the impact of stockouts on the estimator in a clear way. Once we describe this 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.

Let y_i denote the purchase of consumer 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 $j = (0, 1, \ldots, J)$. 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. This leads to the first assumption:

Assumption 1. *(Discrete Choice) Each consumer chooses some product $j \in a_i$ or the outside good $j = 0$.*

For simplicity, we denote the set of possible choices as J, the powerset of $\{0, 1 \ldots, J\}$ 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 jth position if the product is available and zero otherwise.

Without making any parametric assumptions we can write down a multinomial likelihood of seeing an individual choose product j as:[5]

[5]While this seems like a parametric assumption, it should be clear it is without any loss of generality that individual categorical choices follow a multinomial distribution. It's not until we attempt to combine observations across individuals that additional assumptions need to be made.

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

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

$$\begin{aligned} L(y_1, \ldots, y_n|\theta) &\propto \prod_{i=1}^{n} \prod_{\forall j \in J} p_{ij}^{y_{ij}} = \prod_{i=1}^{n} p_{ij} \\ l(y_1, \ldots, y_n|\theta) &\propto \sum_{i=1}^{n} \sum_{\forall j \in J} y_{ij} \ln p_{ij} = \sum_{i=1}^{n} \ln p_{ij} \end{aligned}$$

Typically, researchers specify choice probabilities as $p_j(\tilde{x}_i, \theta)$.[6] Thus, an individual's probability of choosing product j depends on some observable $\tilde{x}_i$ as well as the unknown parameters. The $\tilde{x}_i$ might include observed information about the individual consumer, information about the set of available products (a_i), and perhaps information about the situation under which the choice took place, such as time of day or location (x_i). In the IO literature, researchers often parameterize $p(\cdot)$ as a logit form in order to build flexibility into substitution patterns while avoiding estimation of an unrestricted covariance matrix.

2.1 Exchangeability

When combining data from multiple individual observations in a dataset we typically assume that y_i are IID. However, this is a stronger condition than is necessary to construct the joint likelihood function of the whole dataset. Exchangeability is a weaker condition than IID, in which the outcomes of y_i' can influence other outcomes y_i, so long as all of the information Φ_i relevant to the likelihood contribution of y_i is conditioned on.[7]

Assumption 2. *(Exchangeability of Consumers) Conditional on the information set Φ_i, individual consumers can be re-ordered.*

[6] This is the probability of choosing j *conditional* on $\tilde{x}, \theta$.

[7] This is a trivial result when we consider the space of the (latent) utilities. Even in a relatively complicated random coefficients utility maximization framework such as Berry, Levinsohn, and Pakes (1995), consumers each have an IID draw of ν_{ik} as their random taste for some characteristic and ε_{ij} as their horizontal preference for a particular product. All of these models assume that (ν_i, ε_i) are jointly IID. It is easy to show that IID in latent utilities implies our exchangeability assumption on the choice probabilities, since all of that randomness is integrated out when choice probabilities are computed.

We say that the sequence $((y_1,\Phi_1),(y_2,\Phi_2),\ldots,(y_n,\Phi_n))$ is exchangeable IFF:

$$l\left((y_1,\Phi_1),(y_2,\Phi_2),\ldots,(y_n,\Phi_n)|\theta\right) \quad = \quad l\left(\rho\left((y_1,\Phi_1),(y_2,\Phi_2),\ldots,(y_n,\Phi_n)|\theta\right)\right)$$

holds for any arbitrary permutation operator ρ where $l(\cdot)$ is the likelihood function.

This tells us that once we condition on the relevant Φ_i the likelihood is invariant to the ordering of the observations. The precise form of Φ_i will depend on the particular choice for the likelihood. In the case of a multinomial likelihood and a logit form for $p(\cdot)$, as long as $\Phi_i = (a_i, x_i)$ are observed for each individual observation y_i, then individuals are exchangeable. (That is, the ordering of the i does not affect the likelihood).

It is important to understand the difference between the IID assumption and the exchangeability assumption. For example, in the model we consider, 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 exchangeable in the logit model since choice probabilities depend only on previous sales through $\Phi_i = (a_i, x_i)$. Hence if we know Φ_i and can condition on it, then the ordering of the y_i's no longer affects the likelihood. An example that violates the exchangeability assumption is the presence of pent-up demand following a stockout. In this case a consumer's choice probability for a good would be higher in periods following a stockout. Unless this was captured in the x_i, then the likelihood function would depend on the order in which the y_i were observed.

There are many ways to specify Φ: the entire purchase history of previous individuals prior to a consumer's decision, the full inventory levels of products, etc. The relevant Φ is the set of information for which the likelihood is still exchangeable. For the logit family, this is $\Phi_i = (a_i, x_i)$. Thus even if we observed the full inventory of a vending machine, it would contain no information in terms of the likelihood that was not already captured by a_i because consumers only purchase one unit under the discrete choice assumption.

Exchangeability requires that the population of consumer preferences we sample from cannot change within a level of aggregation, denoted as t. Without this assumption, we cannot tell apart stock-out events from atypical consumers. If this were the case, we could only make inferences about the overall mixture, not its components. For example, if we observed data on sales between 4pm and 8pm, and at 5pm the population of consumers changes, 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.[8] This sort of heterogeneity can be addressed in our approach across periods of observation, but not within a single period of observation. 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.[9]

[8] Previous studies have relied on annual or quarterly data, for which short-term heterogeneity in the population also gets "averaged out" in the overall distribution of consumer preferences.

[9] One method for accommodating latent types is to write down a mixture form for choice probabilities which integrates out the latent variable (this is how we deal with random coefficients in our typical estimators). We describe this approach in the technical appendix.

2.2 Aggregation

We often observe only aggregate data for some period t, denoted $(y_{jt}, \tilde{x}_t) = (\sum_{i \in t} y_{it}, \tilde{x}_{it} = \tilde{x}_t)$.[10] If we believe that within a period t, consumers are *exchangeable* then we can consider aggregate data without loss. In order to do this we require that $\tilde{x}_{it}$ are fully observed and $\tilde{x}_{it} = \tilde{x}_t$ for $i \in t$, or that within our level of aggregation the observables are fixed. There is nothing thus far that requires our periods be contiguous in geography or time, only that they have the same $\tilde{x}_t$.[11] In the case of the logit, this is the same as assuming that for all consumers $i \in t$, $\Phi_i = \Phi_t$ or $(a_i, x_i) = (a_t, x_t)$. This allows us to consider aggregate data is if it were individual purchase data without loss and write the log-likelihood as:

$$l(y_t|\theta, a_t, x_t) = \sum_{j \in a_t} y_{jt} \log p_j(x_t, a_t, \theta) = \sum_{j \in a_t} \sum_{i \in t} y_{it} \log p_j(x_t, a_t, \theta)$$

For simplicity, let $\mathbf{y_t} = [y_{0t}, y_{1t}, y_{2t}, \ldots, y_{Jt}]$. Then for each market, the data provide information on $(\mathbf{y_t}, a_t, x_t)$. By using Assumptions 1 and 2 we can consider the probability that a consumer in market t purchases product j as a function of the set of available products, the exogenous variables, and some unknown parameters θ. This probability is given by

$$p_{jt} = p_j(\theta, a_t, x_t) \tag{1}$$

The key implication of assumptions 1 and 2 is that p_{jt} is constant within a period and does not depend on the realizations of other consumers' choices y_{ijt}. Another immediate implication is that we can reorder the unobserved purchase decisions of individual consumers within a period t. Now, we apply assumption 2 again to write the likelihood function as a multinomial with parameters $n = \sum_{j=0}^{J} y_{jt}$, and $p = [p_{1t}, p_{2t}, \ldots]$

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

Thus $f(\cdot)$ defines a relative measure of how likely it is that we saw the observed data $\mathbf{y_t}$ given the parameter θ. An important simplification arises from the fact that the combinatorial term $C(\mathbf{y_t})$ depends only on the data, and does not vary with the parameter θ. We add a third assumption that is also quite standard in this literature (and generally follows from assuming a utility model).

[10] More formally $\mathbf{y_{jt}} = \sum_{i \in t} y_{ijt} \cdot e_j$ where e_j is the unit vector with 1 in the jth position

[11] We may want geographic or temporal features to enter the $\tilde{x}_t$ though we need to be explicit about it.

Assumption 3. *(Identification) Each period t is exchangeable with respect to other periods, such that for all t, $p_j(\theta, a_t, x_t)$ is a function of the set of available products and some exogenous variables (a_t, x_t) and is known up to a finite dimensional parameter θ.*

Theorem 1. *Taking assumptions 1, 2 and 3, we can define a function S, which takes the vector of observed periods $(\mathbf{y_t}, \mathbf{x_t}, \mathbf{a_t})$ and converts them into their minimal sufficient statistic representation $(\mathbf{y_{t'}}, \mathbf{x_{t'}}, \mathbf{a_{t'}})$, such that:*

$$S\left((\mathbf{y_t}, \mathbf{x_t}, \mathbf{a_t})\right) = (\mathbf{y_{t'}}, \mathbf{x_{t'}}, \mathbf{a_{t'}}) \tag{3}$$

Proof.

$$\begin{aligned} l(\mathbf{y}|\theta, \mathbf{a}, \mathbf{x}) = l(y_1, \ldots, y_T|\theta, \mathbf{a}, \mathbf{x}) &= \sum_t \sum_{j \in a_t} y_{jt} \log p_j(x_t, a_t, \theta) \\ &= \sum_t \sum_{j \in a_t} \sum_{i \in t} y_{ijt} \log p_j(x_i, a_i, \theta) \\ &= \sum_{t'} \sum_{j \in a_{t'}} \sum_{i:(x_i, a_i) = (x_{t'} a_{t'})} y_{ijt} \log p_j(x_i, a_i, \theta) \\ &= \sum_{t'} \sum_{j \in a_{t'}} \sum_{i \in t'} y_{ijt} \log p_j(x_{t'}, a_{t'}, \theta) \\ &= \sum_{t'} \sum_{j \in a_{t'}} y_{jt'} \log p_j(x_{t'}, a_{t'}, \theta) \end{aligned}$$

□

In words, we observe data (y_t, x_t, a_t) aggregated to some level. As long as (x_t, a_t) are fixed in our level of aggregation, we could act as if we had data on independent consumers. Then we can reconstruct a new level of aggregation t' where we aggregate over all observations with the same (x_t, a_t) pairs without changing the value of the likelihood. This gives us some new "pseudo-(period) observations". Typically we say that $y_{jt'}$ (which is just the sum over all y_{jt} with the same (a, x) values) is a sufficient statistic for the likelihood since it contains all of the information about y_i that we need to evaluate the likelihood function. From now on, we assume that any dataset $(\mathbf{y_i}, \tilde{\mathbf{x}}_\mathbf{i})$ can be reconstructed in this minimal sufficient statistic way, and that those groupings (not the original aggregation in the data) are subscripted by t rather than t'.

This leads to the following corollary:

Corollary to Theorem 1. *Since the likelihood is additively separable in the sufficient statistics $S((y_t, x_t, a_t))$, the sums $S((y_t, x_t, a_t))$ can be broken up in an arbitrary way, including* *one sale at a time, as it will not affect the likelihood so long as $S(\cdot)$ assigns sales to the same (a, x) regime.*

Thus, Theorem 1 allows us to reduce the effective number of periods t that we consider. This might help reduce the effective size of our dataset and simplify computation. Conversely, even if we consider the dataset individual by individual, the likelihood function is exactly the same as if we had aggregated over the relevant (a, x) periods. Thus, if we do not observe any variation across (a, x) we essentially only have one multinomial observation. We discuss this in more detail in the section on identification, but the "pseudo-observations" in the sufficient statistic representation and the variation in (a, x) will be what determines identification of θ. If there is not sufficient observable heterogeneity in (a, x) identification becomes a problem (one which cannot be fixed by adding unobservable heterogeneity).

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) and relax the assumption that a_t (the set of available products) is constant across a time period. Instead suppose a stockout occurs in the middle of a period t. Since inventory is observed, the "period" can be divided into two smaller periods of constant availability (before and after the stockout) which we denote (a_s, a_t).

We now know which sales to assign to the pre-stockout regime and which sales to assign to the post-stockout regime (since we observe inventory always). Recalling the likelihood, we see that it remains unchanged when we consider single consumers instead of time periods (Corollary 1).

$$
\begin{aligned}
l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \left(\sum_{\forall j \in a_s} \ln p_j(\theta, a_s, x_t) \sum_{\forall i:(a_{it}, x_{it})=(a_s, x_t)} y_{ji} + \sum_{\forall j \in a_t} \ln p_j(\theta, a_t, x_t) \sum_{\forall i:(a_{it}, x_{it})=(a_t, x_t)} y_{ji} \right) \\
l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \left(\sum_{\forall (a,x)} \sum_{\forall j \in a} \ln p_j(\theta, a, x) \sum_{\forall i:(a_i, x_i)=(a,x)} y_{ji} \right) \\
&= \sum_{\forall t} \left(\sum_{\forall (a,x)} \sum_{\forall j \in a} y_{j,(a,x)} \ln p_j(\theta, a, x) \right) \qquad (4)
\end{aligned}
$$

3.2 Periodic Inventory/ Latent Choice Set Heterogeneity

In many market settings, firms only observe inventories periodically. This presents additional challenges when investigating stock-out events, because availability is known only at the beginning and the end of the period in which a stockout takes place. 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:

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

because $\mathbf{a}$ is not fully observable, so instead we work with the expectation and solve:

$$\hat{\theta} = \arg\max_{\theta} E_{\mathbf{a}} l(\mathbf{y}, \mathbf{a}|\mathbf{a_{obs}}, \mathbf{x}, \theta) \quad = \quad \sum_{\forall \mathbf{a}'} l(\mathbf{y}|\mathbf{a}, \mathbf{x}, \theta) g(\mathbf{a}'|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta). \tag{5}$$

We define $g(\mathbf{a}'|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta)$ as the probability of the sequence of choice sets denoted by $\mathbf{a}$, so that $Pr(\mathbf{a} = \mathbf{a}') = Pr([a_1 = a'_1, a_2, = a'_2, \ldots])$. The expectation in (5) is a summation over all possible availability sets a_i.

The standard thing to do is to partition $\mathbf{a}$ into fully observable and partially observable pieces, so that for each individual sale i, a_i is either known or it isn't. We can then divide the entire dataset into two subsets T_{mis} and T_{obs} and either $i \in T_{obs}$ or $i \in T_{mis}$. The convention is to define $(\mathbf{y_{obs}}, \mathbf{x_{obs}}, \mathbf{a_{obs}})$ as the vector of fully observed data (ie: $\forall i \in T_{obs}$) and $(\mathbf{y_{mis}}, \mathbf{x_{mis}}, \mathbf{a_{mis}})$ as its complement (ie: $\forall i \in T_{mis}$). It should be clear that we're using T_{obs} and T_{mis} because we can easily do the same thing for aggregate data t instead of just individual data without any additional difficulty.[12]

The problem is that evaluating this expectation is difficult. For one, it involves the joint distribution of the a_i's, $g(\mathbf{a}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta)$, which we haven't specified. The other is that this is the expectation of a J-vector and support of $\mathbf{a}$ is potentially large.

$$\begin{aligned}
l(\mathbf{y}|a_{obs}, \mathbf{x}, \theta) &= \sum_{\forall \mathbf{a}'} l(\mathbf{y}|\mathbf{a}, \mathbf{x}, \theta) g(\mathbf{a}'|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) && (6) \\
&= \sum_{i=1}^{n} \sum_{\forall a'_i} l(y_i|a'_i, x_i, \theta) g_i(a'_i|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) && (7) \\
&= \sum_{i \in T_{obs}} l(y_i|a_i, x_i, \theta) + \sum_{i \in T_{mis}} \sum_{\forall a'_i} l(y_i|a_i, x_i, \theta) g_i(a'_i|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) && (8) \\
&= l(\mathbf{y_{obs}}|\mathbf{a_{obs}}, \mathbf{x_{obs}}, \theta) + \sum_{\forall a'_i} l(\mathbf{y_{mis}}|\mathbf{a'_{mis}}, \mathbf{x_{mis}}, \theta) g(\mathbf{a'_{mis}}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) && (9)
\end{aligned}$$

By splitting the likelihood into fully observed and partially observed components we can see that we need only evaluate a single dimensional expectation $n_{mis} = \dim(T_{mis})$ times.

[12]The other technical point here that should be made clear is that the only thing that is ever unobserved is the value of a_t. We assume we still perfectly observe (y_t, x_t) just like we did before.

However, this expectation depends on the distribution $g(\cdot)$, which is, at this point, not yet specified.

3.3 The Case of Stockouts

Stockouts represent an important special case of unobservable choice set heterogeneity, in which the demand model specifies the distribution of $g(\cdot)$ for us. Recall that $g(\cdot)$ specifies the proportion of consumers that faced choice set a_l out of possible choice sets A_t in a given aggregate observation.

In the stockout case, the change in choice set is endogenous. That is, a stockout occurs only when y_{kt} exceeds its capacity ω_{kt}. If we limit ourselves to a single stockout for the sake of exposition, and there are M_t consumers in a particular aggregate observation, then we can ask, how many of the M_t consumers saw choice set a and how many saw choice set a'. Alternatively, we can ask "How many consumers were required in order to sell ω_{kt} units of product k under availability set a?"

This is exactly the definition of the negative binomial distribution, which describes the number of trials m until ω_{kt} successes are observed. Alternatively, it may be formulated as the number of failures r until ω_{kt} successes are observed, where $r + \omega_{kt} = m$. In the case of a stockout with periodic data, we also know that the stockout happened before M_t consumers arrived. In other words, we have a negative binomial, conditional on the fact that $r + \omega_{kt} \leq M_t$.[13]

The definition of the negative binomial is:[14]

$$\begin{aligned}
Pr(z = m) &\sim NegBin(\omega_{kt}, p) \\
f(m, \omega_{kt}, p) &= \frac{(\omega_{kt} + r - 1)!}{m!(\omega_{kt} - 1)!} p^{\omega}_{kt} (1-p)^r
\end{aligned}$$

And the conditional negative binomial:

$$\begin{aligned}
Pr(z = m | z \leq M) &\sim \frac{NegBin(\omega_{kt}, p)}{NegBinCDF(M, \omega_{kt}, p)} \\
h(m', \omega_{kt}, p) &= \frac{f(m', \omega_{kt}, p)}{\sum_{m=1}^{M} f(m, \omega_{kt}, p)}
\end{aligned}$$

For the stockout case where product k stocks out we have that:

[13] This negative binomial is a derived distribution from a multinomial or binomial for waiting times. It should not be confused with negative binomial regression (often used for count data), which is just an overdispersed Poisson model.

[14] This definition is presented in terms of the more familiar factorial. It is often useful to consider the non-integer generalization, the gamma function $\Gamma(x + 1) = x!$.

$$\begin{aligned}\alpha_t(m) &= \frac{r+\omega_{kt}}{M_t} = \frac{m}{M_t} \\ Pr\left(\alpha(m)=\alpha'|\alpha\leq 1\right) &\sim \frac{NegBin(m,y_{kt},p_k(a,x_t,\theta))}{NegBinCDF(M,y_{kt},p_k(a,x_t,\theta))}\end{aligned}$$

In other words, α_t is the fraction of consumers in market t arriving before the stockout, and $1-\alpha_t$ is the fraction of consumers facing a stockout. Therefore α takes on a discrete set of values from 0 to 1 and has a probability mass function (p.m.f) given by the conditional negative binomial distribution $g(\cdot) = Pr\left(\alpha|\alpha\leq 1\right)$. In the case of multiple stockouts, $g(\cdot)$ has a conditional negative multinomial distribution. We show this result in sections A.1 and A.2 of the appendix.

Now we can evaluate the mixture likelihood (with availability a before the stockout and a' after):

$$l_t(y_t,A_t,x_t,\theta) = \sum_{j\in A_t} y_{jt} \sum_{\alpha_t:m\leq M_t} \ln\left(\alpha_t\cdot p_j(x_t,a,\theta)+(1-\alpha_t)\cdot p_j(x_t,a',\theta)\right) g(\alpha_t|y_t,x_t,M_t)$$

The distribution $g(\cdot)$ depends only on data from its own observation t, and doesn't depend on realizations of a_t from other observations. This is an implication of the model for demand under stockouts, not an assumption. Therefore, we can evaluate expectations by evaluating the density $g(\cdot)$ over the finite support of $\alpha(m)$, which should not prove too difficult, since it is just two sums over $J\times M_t$ elements (rather than some high dimensional expectation). When M_t gets particularly large (or a large number of stockouts happen at once) this could still present some computational challenges. One might consider approximating the expectation with Monte Carlo or quadrature techniques as detailed in section A.3 of the appendix.

Additionally, the likelihood (9) must still be evaluated at each guess of θ, which can be time-consuming in practice. Furthermore, the dependence of the integral on $g(\cdot)$ also affects the analytic gradients since they depend on $g(\cdot)$ too. Thus, instead of directly evaluating this expectation, we specify an augmented model and use the E-M algorithm to facilitate estimation.

3.4 Augmented Model

If we suppose that we had perpetual inventory then we could denote the set of availability regimes potentially observed in period t as A_t and define an augmented model as:

$$l_t(y_{ta_l},A_t,x_t,\theta) = \sum_{a_l\in A_t}\sum_{j\in a_l} y_{jta_l}\ln p_j(x_t,a_l,\theta)$$

In the augmented model, we know the set A_t, since we observe the availability at the beginning and end of a period. In this case the observed sales for observation t are distributed multinomially (across $J \times \dim(A_t)$ cells).[15] Note that with periodic data collection, in the augmented model, y_{jta_l}'s are not observed, so we cannot evaluate $l_t(y_{ta_l}, A_t, x_t, \theta)$ directly. But, we replace it with a consistent estimator (its expectation). Then the likelihood becomes:

$$l(\mathbf{y}, \mathbf{x}, \mathbf{A}, \theta) = \sum_{t \in T_{obs}} l_t(y_t, a_t, x_t, \theta) + \sum_{t \in T_{mis}} E_{\alpha(\theta')}[l_t(y_{ta_l}, A_t, x_t, \theta)]$$

This is now a likelihood function of only the observables (for a fixed value of θ'). By iterating back and forth between computing this expectation at some value of θ', plugging this in and then maximizing the complete-data likelihood over θ, and updating θ' we can obtain consistent parameter estimates for θ. This is the well known E-M Algorithm of Dempster, Laird, and Rubin (1977). It should be clear that this approach is no different from integrating out α above, except that evaluating this expectation is much easier.

We want to evaluate:

$$\begin{aligned} E_\alpha[l_t(y_{ta_l}, A_t, x_t, \theta)] &= E[\sum_{a_l \in A_t} \sum_{j \in a_l} y_{jta_l} \ln p_j(x_t, a_l, \theta)] \\ &= \sum_{a_l \in A_t} \sum_{j \in a_l} E_\alpha[y_{jta_l}] \ln p_j(x_t, a_l, \theta)] \end{aligned}$$

We only need to find $E_\alpha[y_{jta_al}]$, because no other quantities are random (the function $p(\cdot)$ is fixed once we know a_l). Furthermore, the likelihood is linear in the unobservable y_{jta_l}, so we can evaluate the expectation separately for each t, and this expectation is a univariate integral so long as $g(\alpha|y_t, x_t, \theta)$ does not depend on data from other observations/periods. We already know that this is true for the case of stockouts.

Our objective is then to evaluate $E_\alpha[y_{jta_l}|\mathbf{y}, \mathbf{x}, \mathbf{A_t}, \theta]$, but recall that we already know the distribution of $(y_{jta_l}|y_{jt})$. For the case of a single stockout:

$$\begin{aligned} (y_{jta_l}|y_{jt}, \alpha_t) &\sim Bin(n = y_{jt}, \gamma_{jt}) \\ \gamma_{jt} &= \frac{p_j(x_t, a_l, \theta)\alpha_t}{p_j(x_t, a_k, \theta)\alpha_t + p_j(x_t, a_k, \theta)(1-\alpha_t)} \\ E[(y_{jta_l}|y_{jt}, \alpha_t)] &= y_{jt} \frac{p_j(x_t, a_l, \theta)\alpha_t}{p_j(x_t, a_k, \theta)\alpha_t + p_j(x_t, a_k, \theta)(1-\alpha_t)} \end{aligned}$$

[15] Recall that when we exactly observe the timing of stockouts, we can simple break up the data into $\dim A_t$ observations each with J cells, and estimate the likelihood directly because we observe the sales in each regime (the y_{jta_l}'s). (Restatement of Theorem 1). This avoids the need to calculate the expectation in equation 9 that depends on $g(\cdot)$.

We can evaluate this expectation over α, whose distribution is known for the case of stockouts:

$$E[(y_{jta_l}|y_{jt})] = y_{jt} \sum_{\forall \alpha_t} \frac{p_j(x_t, a_l, \theta)\alpha_t}{p_j(x_t, a_k, \theta)\alpha_t + p_j(x_t, a_k, \theta)(1-\alpha_t)} g(\alpha_t|y_{jt}, x_t, \theta) \tag{10}$$

We can now treat our expected cell counts $E[y_{jta_l}] = \hat{y}_{jta_l}$ as if they were observed data, and can evaluate the complete data likelihood as if there were no missing data problem. It should be made clear that technically the imputed values of y depend on θ, in other words, $\hat{y}_{jta_l}(\theta)$.

This gives us a procedure for computing the ML estimate of θ in the presence of unobservable choice sets. We simply iterate over the following steps until we reach a fixed point:

$$\begin{aligned}
\hat{y}_{jta_l}(\theta^t) &= y_{jt} \sum_{\forall \alpha_t} \frac{p_j(x_t, a_l, \theta^t)\alpha_t}{p_j(x_t, a_k, \theta^t)\alpha_t + p_j(x_t, a_k, \theta^t)(1-\alpha_t)} g(\alpha_t|y_{jt}, x_t, \theta^t) \\
\theta^{t+1} &= \arg\max_{\theta} \sum_{t \in T_{obs}} \sum_{j \in a_l} y_{jt} \ln p_j(x_t, a_t, \theta) + \sum_{t \in T_{mis}} \sum_{a_l \in A_t} \sum_{j \in a_l} \hat{y}_{jta_l}(\theta^t) \ln p_j(x_t, a_l, \theta)
\end{aligned}$$

These are known as the (E-Step) and (M-Step) respectively. When we iterate over these we monotonically increase the likelihood until we obtain the ML estimate of θ that is the same as if we had evaluated the integral from the previous section. There are two advantages of this procedure: computational ease, and the fact that the missing data have a sensible interpretation in this context (which sales occurred before and after the stockout event).

The major computational advantage is that we can use an off-the-shelf procedure for ML multinomial logits once we've imputed the missing data, since the likelihood is exactly the same as the complete data case. In other words, any gradient-based procedures for optimization of the likelihood will still work for the M-Step if they worked for the complete data problem. Furthermore, we only evaluate the expectation after we fully maximize the likelihood over θ. In practice, this means the difference between integrating fewer than 100 times versus integrating several million times. The exact number of E-M iterations depends on how much information is "missing".

3.5 An Alternative Approach

One potential disadvantage of the approach just described is that it relies on an assumption of exchangeability. Stockouts imply a distribution on $g(\cdot)$ 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 like those used here. It is unclear whether exchangeability 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 of their tastes). Our estimator leverages the exchangeability assumption more heavily before, in that we use the fact that the distribution of tastes is fixed before and after the stockout to assign sales to regimes. This in turn helps to define the $g(\cdot)$ distribution separately from the θ vector.

One might wonder whether alternative approaches exist that do not lean so heavily on the exchangeability of consumers. For example, one could consider a fully simulated approach, where, for a guess of θ, we record the initial inventory and number of consumers M_t for each interval, and then simulate individual consumer choices, repeating for each of the M_t consumers in the market.[16] We would repeat this process some large number of times for each t and compute the average purchase probability for each product, denoting it as $\hat{p}_{jt}(\theta)$.[17]

If each consumer's utilities were determined by IID draws of random tastes and idiosyncratic preferences $(\nu_{ik}, \varepsilon_{ij})$) then this would be a far less efficient version of our estimator presented in the previous section. However, if we felt it were an important feature of the market in question, the simulated approach affords us the flexibility to allow the distribution of tastes to vary over time, even in a way that allowed for correlation with stockouts.[18] In fact, this indirect-inference type approach could be utilized for any scenario where we could write down a series of functions that allowed us to simulated consumer purchases. Once we are able to compute the choice probabilities $\hat{p}_{jt}$ from a more complicated simulated procedure, our best bet is probably to do simply evaluate the likelihood (MSL) just as in the previous case. [19]

In the case of latent stockouts, one could match the aggregate sales for each time period conditional on the starting inventory. In contrast, our method breaks up time periods when there are changes in product availability and matches imputed sales to predicted sales. Even considering an infinite number of trials, the alternative approach is not fully efficient like the EM procedure, where the major discrepancy arises because the procedure is not restricted to availability regimes that occurred in the data. In other words, it does not condition on the full set of available data, and instead will average over all possible availability regimes encountered by simulation. Augmenting the procedure to condition on this information would require conditioning on the outcome, which would ruin its simplicity (and introduce additional integrals).

Taking this together with the fact that the exchangeability assumption does not seem too far-fetched in our application, we do not report results from such a method here. However,

[16]Simulating 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, and then (4) update the inventory and number of consumers remaining in the market.

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

[18]We could imagine adapting our EM type procedure to these scenarios. Doing so would likely imply a more complicated functional form for $g(\cdot)$ and (10).

[19]We use some MSM procedure, such as the GMM procedure from Berry, Levinsohn, and Pakes (1995) which deals with endogenous prices, or allows us to incorporate additional restrictions. In general, ML type approaches are efficient and easy to implement. Once we use the starting inventories to simulate "mixed" choice probabilities, we may no longer need to worry so much about the extremely granular nature of high-frequency data, and large-sample method of moments procedures may become feasible.

we provide further details on this type of approach in the appendix (section A.4), 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 in the differentiated products literature to 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, 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 (error-components) nested logit models can be nested in the above framework. For the nested logit, $\mu_{ijt} = \sum_g \sigma_g \zeta_{jg} \nu_{ig}$, where $\zeta_{jg} = 1$ if product j is in category g and 0 otherwise, and ν_{ig} is standard normal. 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 lth characteristic of product j and ν is standard normal. In both models, the unknown parameters are the σ's. This representation makes it clear that the nested logit is a special case of the random coefficients logit.

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.

When using a random coefficients logit form, we estimate via Maximum Simulated Likelihood (MSL) rather than ML. For the multinomial choice model the MSL estimator is easy to define. Begin 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.

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

4.2 Heterogeneity

Thus far, we've done everything conditional on x_t. In one sense, this is useful to show that our result holds for the case of conditional likelihood, but it is also of practical significance to our applied problem. Since periods in retail datasets may be short, it is likely that choice probabilities may vary substantially over periods. 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 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 of Discrete Choice Models

In this section we address non-parametric and parametric identification of the choice probabilities $p_j(\theta, a_t, x_t)$, when the underlying data generating process is multinomial. The goal is not to provide formal identification results, but rather to provide a clear exposition so that the applied researcher can better understand the practical aspects of identification in the discrete choice context.[20].

Recall that we express our sufficient statistics as the vector $\mathbf{S}((\mathbf{y_t}, \mathbf{x_t}, \mathbf{a_t}))$. Denote $q_j(a, x)$ as the sufficient statistic corresponding to 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:

$$\hat{p_j}(a, x) = \frac{q_j(a, x)}{M_{(a,x)}}$$

The variance of $\hat{p_j}(a, x)$ can be written as:

$$\begin{aligned} Var(\hat{p_j}(a, x)) &= Var\left(\frac{1}{M_{a,x}} q_j(a, x)\right) = \left(\frac{1}{M_{a,x}}\right)^2 \cdot M_{a,x} \cdot p_j(a,x)(1 - p_j(a, x)) \\ &= \left(\frac{p_j(a,x)(1 - p_j(a, x))}{M_{a,x}}\right) \approx \left(\frac{p_j(a,x)}{M_{a,x}}\right) \end{aligned}$$

Thus, the variance of nonparametric estimators for $p_j(a, x)$ will go to zero as $M_{a,x} \to \infty$. This variance is typically referred to as "measurement error in the choice probabilities," as we could think about a two-stage procedure where we nonparametrically recover choice probabilities from observed marketshares and then fit a parametric $p(\cdot)$ function to these choice probabilities.[21] It also tells us that the variance as a percentage of the shares de-

[20] Athey and Imbens (2007) provide some related identification results for the fully Bayesian MCMC estimator for these sorts of models. As already discussed, our approach could be computed using such an MCMC approach as well.

[21] For 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 $M_{a,x} \to \infty$. While this might be reasonable for annual data at a national level, it becomes more problematic in the analysis of high-frequency

clines almost uniformly across share sizes.[22] At the same time, unless every (a, x) pair in the domain is observed (and with a substantial number of consumers) the conditional mean (semi-nonparametric) representation of our $p_j(\theta, a_t, x_t)$'s will not be nonparametrically identified.

The problem of nonparametric identification in these models is well known, and the standard starting point is to write down a functional form for either the probabilities or the underlying utility function. The first question a researcher might ask is "What sort of variation do I need to see in the data to identify this function?" Most of the approaches follow the econometric literature beginning with Matzkin (1992). For example, Ackerberg 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) provides 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 our approach is motivated by an experimental or quasi-experimental setting in which we look at variation among available choice sets. We ask the question,"If what we are trying to measure are tastes and substitution across products, what sort of choice-set experiments would we need in order to identify the distribution of tastes?". The underlying intuition is quite simple, if what we're trying to identify is product specific effects, the easiest way to identify these effects should be to conduct an experiment where we remove the product from the choice set, and record the sales.

Our sufficient statistic representation makes this really easy to see, because we can consider a linearized world. Recall our $S(\cdot)$ operator takes a set of data and returns to us the minimal sufficient statistic representation in terms of (a, x) "pseudo-observations". If we observe only a single choice set throughout our sample, then we have only a single observation in our data. However, for the logit form, it is easy to see that from this single observation we could identify the d_j product dummies for every product that was available. In the absence of any random coefficients, all we need is a single observation in which all products are available in order to identify the linear parameters d_j (by matching the average choice probabilities). 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 x_j, 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 rather the ratio of the choice probabilities before and after a stockout.[23]

Intuitively, one can think of the periods of observation in high-frequency datasets as being of three types. In type 1, no products stock out. This is the usual scenario. In type 2, a product stocks out at the very beginning of the period. This creates good variation in

data.

[22]Thus we don't need more observations to correctly estimate smaller shares than we need to estimate larger shares.

[23]We provide this ratio and its derivation in section A.5 of the appendix.

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. Type 3 markets are periods in which a product stocks out sometime during t. Periodic data collection will generate these types of observations in addition to types 1 and 2. Ignoring these observations selects on sales levels, so we need to incorporate them, which is exactly what our missing data procedure does.

Stockout events are useful, particularly when trying to identify product dummies and nonlinear parameters, because they provide linearly independent observations of $q_{a,x}$. For example, if we only ever observed a change from choice set $a \rightarrow a'$ (suppose the only product that ever stocks out is Snickers), then we would have two effective 'observations'. If we observe lots of stockouts and different choice sets, then we have the potential to observe 2^J 'observations'.

In the typical approach (e.g., Berry, Levinsohn, and Pakes (1995) and related literature), the choice set a_t is often not a collection of products, but rather a collection of bundles of characteristics. Thus a_t is not the set of available products $j = 1, \ldots, J$ as it is in our model, but rather the set of available characteristic bundles, including price and product characteristics. Our sufficient statistic representation provides an alternative explanation: if characteristics vary "continuously" and at random, then after applying our sufficient statistic data reduction operator $\mathbf{S}(\cdot)$, we still have the effective number of pseudo-observations going to infinity. Even with an infinite number of pseudo-observations, it is not guaranteed that the observations are linearly independent and all linear and nonlinear parameters are identified. [24]

In practice, much empirical work uses a characteristics-based approach, and relies on changes in stocking decisions, possibly changes in non-price product characteristics, and a finite amount of (possibly small) price variation. As Ackerberg 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 case, we don't observe price variation, and we do not attempt to identify a price coefficient. If we wanted to pin down a price coefficient, we would want variation across prices in addition to changes in the choice set.[25]

In summary, our model presents a different way to interpret variation in choice sets. In

[24]Once again the $q_{a,x}$ representation makes it clear that additional draws from the same choice set do not provide additional observations, but rather get added into $q_{a,x}$, which may improve efficiency of the non-parametric step, but does not allow us to identify additional parameters.

[25]If instrumenting for price, one would need to either add additional distributional assumptions to the ML problem, or use a GMM procedure. Draganska and Jain (2004) develops a 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 granular datasets such as ours. 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 where we have individual-level data. Finally, Allenby, Chen, and Yang (2003) provide a Bayesian estimator when IVs are required.

our context a_t doesn't vary with long-term product mix, or potentially endogenous pricing decisions, but rather as products stock out. Additionally, we don't need to worry about this sort of variation being endogenous, because firms take changes in consumer's choice sets as given, and so do consumers. This might not seem obvious at first, but because choice sets are realizations of stochastic choices of consumers, and consumers' choices depend only on the set of available products, stockouts are random events.[26] In fact they are fully described by our $g(a_i|\cdot)$ distribution. While firms can restock the machine (or even change the product mix to prevent future stockouts) and these actions might change the probability that a particular consumer faces a stockout, once we condition on that probability, the occurrence of any particular stockout is a random occurrence and thus exogenous to the model. Finally, one of the most common applications of these models is to predict substitution probabilities. Perhaps the best way to predict substitution probabilities is to observe them. Stockouts provide not only a chance to observe substitution probabilities, but also an opportunity to observe them repeatedly and across different dimensions than previous approaches have been able to observe.[27]

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.[28]

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

[26]Recall that the primitive is preferences, not choices, and so as long as preferences do not depend on choice sets, choice set variation is exogenous to the model because it depends on stochastic outcomes of consumer arrivals.

[27]For a discussion of the benefit of choice subsets, see Zeithammer and Lenk (2006).

[28]In 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.

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, 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 items[29] or have insubstantial sales (usually less than a dozen total over all machines).[30] 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).[31] Finally, we combined two different versions of three products.[32] The 44 products in the final dataset are listed in Table 1.

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

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

[31] Misc 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.

[32] These 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.

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 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 Oreo cookies, 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 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. 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 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.[33] 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 similar, and so we accommodate heterogeneity through M.[34] We divide machines into three

[33] The 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.

[34] Thus, it is assumed that the choice probabilities (the $p_j(\cdot)$ functions) are stable across markets as

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 in a weekday we have 250 consumers, and on weekend days, about 75 consumers, for a total of 1400 consumers per week on average for at a machine in the middle tier. 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.[35] 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–for example, say we allowed choice probabilities to differ freely for each machine–the fewer consumers we have for any given availability set, because now we can only use the observations generated by one machine.

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.[36]

Table 4 reports the number of choice sets, log likelihood, market size, and nonlinear parameters from estimation of nested-logit and random-coefficients specifications under each of the three treatments of stock-out events.[37] 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).[38] 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).

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 consumers over the two semesters. The correlation within nest is approximately $(1-\lambda)$, or

discussed in the estimation section.

[35]We use some additional conditions to prevent sales from exceeding the marketsize, particularly in very short periods, but in general these are not binding.

[36]All daily-level results are available upon request from the authors.

[37]All 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).

[38]We 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.

0.48 under an assumption of full availability. The second column (the 'Ignore' model) uses 2649 unique choice sets. 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 fit of the model (as measured by the log-likelihood divided by market size) improves dramatically–but much of this improved fit comes from the fact that we have eliminated many of the observations that are the most difficult to fit. 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. 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 and fit statistic 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.[39]

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 candy $(1-\lambda)=0.54$, and lowest for pastry $(1-\lambda)=0.09$.

The third panel of table 4 reports the same set of estimates using a random-coefficients logit model for demand in which random coefficients are estimated for each of three observable product characteristics: fat, sodium, and sugar.[40] The random-coefficients specification has a higher likelihood than either of the nested-logit specifications under an assumption of full availability. However, the likelihood under both of the nested-logit specifications exceed the random-coefficients specification under the Ignore and EM-corrected models, indicat-

[39] One 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 $g(\cdot)$ in equation 9, as detailed in section 3. Substituting arbitrary distributions for $g(\cdot)$ will not give consistent estimates of θ, because the demand model implies a specific $g(\cdot)$ 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.6 of the appendix.

[40] We 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)).

ing that the observable product characteristics may be less useful than product categories for organizing correlation structures in this market. 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).[41]

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 most cases. When the predicted best substitutes differ, we can 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.[42] 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 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 Babyruth, 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 243

[41] We 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.

[42] We 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.

or 227 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.[43] 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.[44]

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 (with one exception of chocolate bars under the nested-logit model). Within any given category, this reflects lower estimated correlation in tastes (the exception being the chocolate category in the nested-logit specification). However, this effect 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., Nutter Butter Bites sell 2.42 additional units according to the Full Availability model, but only 1.41 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., 4.89 units of M&M Peanut are sold according to the EM-corrected model, compared to 1.05 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, since 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.

[43]We 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.

[44]As 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.

In the lower panel of table 7, 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, 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 staying inside (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 - 32 percent vs. 30 - 37 percent for Full Availability). Gross profit is thus lower under the EM-corrected model, with a loss from the stockouts of roughly \$37 to \$40 compared to \$31 to \$36 for the Full Availability model. In percentage terms, this difference is 16.6% in the nested-logit specification, and 11.6% in the random-coefficients specification. Thus, profit losses from stockouts are roughly 12 - 16 percent larger than the standard Full Availability model would have us believe. For an industry with profit margins of less than 4%, this is a significant difference.[45]

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 1 shows the median change in the sales of substitutes by rank across all 35 products.[46] 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.

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, 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

[45]Companies 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).

[46]The set of 35 individual graphs are available upon request from the authors.

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.

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. As these data become 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.

Product	Category	% SO (Lower)	% SO (Upper)	p	c	Share	Avg D Sales	No. Mach.
PopTart	Pastry	4.81	5.87	1.00	0.35	3.71	0.90	54
Choc Donuts	Pastry	15.09	17.51	1.00	0.46	2.95	0.81	53
Ding Dong	Pastry	12.93	14.99	1.00	0.46	2.79	0.73	54
Banana Nut Muffin	Pastry	6.85	7.98	1.00	0.46	2.73	0.66	54
Rice Krispies	Pastry	1.14	1.38	1.00	0.31	1.99	0.50	54
Pastry	Pastry	8.80	10.48	1.00	0.46	0.90	1.03	23
Gma Oatmeal Raisin	Cookie	1.20	1.49	0.75	0.23	2.74	0.69	52
Chips Ahoy	Cookie	0.70	0.92	0.75	0.25	2.46	0.60	53
Nutter Butter Bites	Cookie	0.16	0.20	0.75	0.26	1.75	0.45	51
Knotts Raspberry Cookie	Cookie	0.35	0.45	0.75	0.19	1.72	0.43	52
Gma Choc Chip	Cookie	0.31	0.36	0.75	0.22	1.53	0.78	52
Gma Mini Cookie	Cookie	1.41	1.73	0.75	0.21	0.80	0.64	52
Gma Caramel Choc Chip	Cookie	2.52	2.93	0.75	0.23	0.52	0.68	52
Rold Gold	Chips	6.83	8.18	0.90	0.27	4.02	0.97	54
Sunchip Harvest	Chips	5.36	6.44	0.90	0.27	3.84	0.93	54
Dorito Nacho	Chips	1.58	1.89	0.90	0.27	3.36	0.81	54
Cheeto Crunchy	Chips	3.91	4.64	0.90	0.27	3.35	0.81	54
Gardetto Snackens	Chips	0.72	0.78	0.75	0.26	3.20	1.20	54
Ruffles Cheddar	Chips	1.59	1.87	0.90	0.27	2.81	0.68	54
Fritos	Chips	2.14	2.41	0.90	0.27	1.86	0.45	54
Lays Potato Chip	Chips	2.50	2.85	0.90	0.27	1.69	0.41	54
Munchies Hot	Chips	0.89	1.00	0.75	0.25	1.42	0.84	51
Misc Chips 2	Chips	1.93	2.17	0.90	0.28	1.30	0.34	54
Munchies	Chips	2.91	3.25	0.90	0.25	1.24	0.53	53
Misc Chips 1	Chips	2.46	2.69	0.90	0.26	1.13	0.57	53
Dorito Guacamole	Chips	0.54	0.63	0.90	0.28	0.94	0.47	53
Snickers	Chocolate	0.65	0.87	0.75	0.33	8.35	2.01	54
Twix	Chocolate	0.52	0.67	0.75	0.33	6.21	1.49	54
M&M Peanut	Chocolate	1.40	1.75	0.75	0.33	4.69	1.13	54
Reese's Cup	Chocolate	0.62	0.73	0.75	0.33	2.37	0.57	54
Kit Kat	Chocolate	0.48	0.60	0.75	0.33	2.16	0.52	54
Caramel Crunch	Chocolate	0.40	0.49	0.75	0.33	2.14	0.52	54
M&M	Chocolate	1.02	1.12	0.75	0.33	1.69	0.60	54
Hershey Almond	Chocolate	0.30	0.36	0.75	0.33	1.65	0.40	54
Babyruth	Chocolate	3.44	3.81	0.75	0.28	0.47	0.35	53
Starburst	Candy	0.70	0.90	0.75	0.33	3.18	0.96	54
Kar Nut Sweet/Salt	Candy	1.21	1.44	0.75	0.22	2.85	0.69	54
Snackwell	Candy	0.39	0.44	0.75	0.28	1.66	0.41	54
Skittles	Candy	0.39	0.53	0.75	0.34	1.62	0.82	54
Payday	Candy	0.00	0.00	0.75	0.33	1.15	0.53	54
Oreo	Candy	0.10	0.12	0.75	0.22	1.01	0.25	54
Peanuts	Candy	0.37	0.42	0.75	0.26	0.86	0.45	51
Peter Pan (Crck)	Candy	0.13	0.14	0.75	0.12	0.81	0.42	49
Hot Tamales	Candy	3.38	3.98	0.75	0.27	0.40	0.49	54

Table 1: Summary of Products and Markups

	(1)	(2)	(3)
# Products Stocked Out	-0.895***	-0.437***	-0.408***
	(0.017)	(0.021)	(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***	7.727***	7.743***
	(0.049)	(0.050)	(0.050)
Observations	111195	111195	111195
R^2	0.0238	0.0349	0.0367

Table 2: Regression of Profit on Stock-Out Variables

	OLS (1)	OLS (2)	OLS (3)	Probit (1)	Probit (2)	Probit (3)
Starting Inven	-0.0168***	-0.00872***	-0.0101***	-0.0142***	-0.00681***	-0.00893***
	(0.00025)	(0.00042)	(0.00055)	(0.00022)	(0.00035)	(0.00036)
Hours	0.00203***	0.00217***	0.00278***	0.00169***	0.00174***	0.00212***
	(0.00011)	(0.00010)	(0.00011)	(0.000092)	(0.000081)	(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

Table 3: Regression of Stockout Rates on Starting Inventory

	Full	Ignore	EM
Single Nesting Parameter (λ)			
Category	0.525	1.075	0.768
Neg LL (millions)	2.280	1.907	2.226
LL/M	0.394	0.358	0.385
Category Specific Nesting Parameter (λ)			
Pastry	1.550	1.131	0.912
Cookie	0.172	0.677	0.521
Chips	0.168	1.465	0.826
Chocolate	0.889	0.767	0.487
Candy	0.184	0.492	0.467
Neg LL (millions)	2.279	1.907	2.226
LL/M	0.394	0.358	0.385
Random Coefficients			
Fat	0.566	0.000	0.306
Salt	2.851	0.000	2.523
Sugar	5.638	0.000	4.822
Neg LL (millions)	2.280	1.907	2.267
LL/M	0.394	0.358	0.392
Choice Sets	238	2649	3966
Marketsize (millions)	5.787	5.320	5.787

Table 4: Non-linear Parameter Estimates

Product	Nested	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	PopTart
Chips Ahoy	Gma Choc Chip	Snickers
Nutter Butter Bites	Gma Choc Chip	Snickers
Knotts Raspberry Cookies	Gma Choc Chip	Banana Nut Muffin
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	M&M	Banana Nut Muffin
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

Table 5: Best Substitutes, Top 35 Products

	Category Specific Nested Logit				Random Coefficients Logit			
	Full	Ignore	EM	(EM-Full)	Full	Ignore	EM	(EM-Full)
PopTart	9.02	8.86	9.51	0.49	9.72	9.78	10.24	0.52
Choc Donuts	8.69	9.78	10.54	1.85	9.44	10.85	11.63	2.18
Ding Dong	7.67	8.05	8.81	1.14	8.20	8.90	9.55	1.35
Banana Nut Muffin	6.63	6.52	7.06	0.44	7.27	7.19	7.68	0.41
Rice Krispies	4.95	4.54	4.93	-0.02	5.52	4.99	5.56	0.04
Gma Oatmeal Raisin	6.40	6.16	6.79	0.40	7.61	7.01	7.59	-0.03
Chips Ahoy	5.57	5.25	5.85	0.28	6.70	6.02	6.73	0.03
Nutter Butter Bites	4.08	3.85	4.24	0.16	4.88	4.39	4.86	-0.02
Knott's Raspberry Cookie	3.99	3.81	4.15	0.17	4.73	4.33	4.71	-0.02
Gma Choc Chip	8.48	7.19	8.15	-0.33	8.60	7.70	8.53	-0.08
Rold Gold	9.83	9.53	10.60	0.77	11.18	10.61	12.08	0.90
Sunchip Harvest	9.37	8.86	9.88	0.51	10.41	9.85	11.10	0.69
Dorito Nacho	8.17	7.38	8.14	-0.03	9.08	8.18	9.16	0.08
Cheeto Crunchy	8.16	7.55	8.35	0.19	9.15	8.38	9.42	0.26
Ruffles Cheddar	6.87	6.21	6.84	-0.03	7.66	6.88	7.71	0.05
Fritos	4.53	4.03	4.52	-0.01	5.04	4.47	5.09	0.05
Lays Potato Chip	4.13	3.77	4.14	0.02	4.59	4.16	4.67	0.08
Munchies Hot	7.76	7.37	7.84	0.08	8.79	8.02	8.84	0.05
Misc Chips 2	3.40	3.13	3.48	0.08	3.81	3.46	3.92	0.12
Munchies	5.33	4.78	5.40	0.07	5.86	5.33	6.06	0.20
Dorito Guacamole	4.68	4.07	4.57	-0.11	5.12	4.55	5.13	0.01
Snickers	20.04	18.37	20.07	0.03	21.91	20.63	21.92	0.02
Twix	14.92	13.39	14.87	-0.05	16.33	15.04	16.30	-0.03
M&M Peanut	11.29	10.36	11.41	0.12	12.36	11.63	12.55	0.19
Reese's Cup	5.69	5.14	5.66	-0.03	6.30	5.77	6.29	-0.01
Kit Kat	5.20	4.68	5.17	-0.03	5.62	5.26	5.61	-0.01
Caramel Crunch	5.17	4.71	5.13	-0.04	5.69	5.30	5.68	-0.02
M&M	5.69	5.25	5.78	0.09	6.27	5.84	6.29	0.01
Hershey Almond	3.97	3.61	3.94	-0.03	4.40	4.06	4.39	-0.01
Starburst	8.47	7.92	8.78	0.31	9.57	9.19	9.65	0.07
Kar Nut Sweet/Salt	6.29	5.91	6.57	0.27	7.60	6.98	7.64	0.04
Snackwell	3.71	3.51	3.90	0.19	4.52	4.16	4.57	0.05
Skittles	7.97	7.26	8.06	0.10	9.01	8.01	8.91	-0.09
Oreo	2.27	2.12	2.37	0.10	2.77	2.52	2.77	0.00
Peanuts	4.45	4.08	4.47	0.02	4.88	4.46	4.90	0.02
Total	242.83	226.97	250.00	7.17	243.59	228.55	250.00	6.41

Table 6: Predicted Weekly Sales

	Category Specific Nested Logit			Random Coefficients Logit		
	Full	Ignore	EM	Full	Ignore	EM
PopTart	-9.02	-8.86	-9.51	-8.75	-8.80	-9.21
Choc Donuts	-8.69	-9.78	-10.54	-8.50	-9.77	-10.47
Ding Dong	-2.23	-0.54	0.69	3.11	0.17	2.97
Banana Nut Muffin	-1.93	-0.44	0.55	4.02	0.14	3.60
Rice Krispies	-1.44	-0.30	0.39	0.56	0.10	0.49
Gma Oatmeal Raisin	-6.40	-6.16	-6.79	-6.85	-6.31	-6.83
Chips Ahoy	-5.57	-5.25	-5.85	-6.03	-5.42	-6.06
Nutter Butter Bites	2.42	0.87	1.41	0.56	0.09	0.48
Knott's Raspberry Cookie	2.36	0.86	1.39	0.94	0.08	0.77
Gma Choc Chip	5.02	1.63	2.72	1.83	0.15	1.49
Rold Gold	-9.83	-9.53	-10.60	-10.06	-9.55	-10.88
Sunchip Harvest	-9.37	-8.86	-9.88	-9.37	-8.86	-9.99
Dorito Nacho	2.55	-0.90	0.61	0.44	0.16	0.41
Cheeto Crunchy	2.55	-0.92	0.63	0.83	0.16	0.72
Ruffles Cheddar	2.15	-0.76	0.52	0.44	0.13	0.39
Fritos	1.41	-0.49	0.34	0.20	0.09	0.19
Lays Potato Chip	1.29	-0.46	0.31	0.19	0.08	0.18
Munchies Hot	2.42	-0.90	0.59	0.47	0.16	0.42
Misc Chips 2	1.06	-0.38	0.26	0.15	0.07	0.15
Munchies	1.66	-0.58	0.41	0.42	0.10	0.38
Dorito Guacamole	1.46	-0.50	0.35	0.16	0.09	0.15
Snickers	-20.04	-18.37	-20.07	-19.72	-18.57	-19.73
Twix	-14.92	-13.39	-14.87	-14.70	-13.54	-14.68
M&M Peanut	1.05	1.98	4.89	2.77	0.23	2.32
Reese's Cup	0.53	0.98	2.42	1.21	0.11	0.99
Kit Kat	0.48	0.89	2.22	1.67	0.10	1.38
Caramel Crunch	0.48	0.90	2.20	1.18	0.10	0.97
M&M	0.53	1.00	2.48	1.94	0.11	1.61
Hershey Almond	0.37	0.69	1.69	0.66	0.08	0.54
Starburst	-8.47	-7.92	-8.78	-8.62	-8.28	-8.68
Kar Nut Sweet/Salt	-6.29	-5.91	-6.57	-6.84	-6.28	-6.88
Snackwell	2.38	1.33	1.55	0.76	0.08	0.63
Skittles	5.11	2.76	3.20	4.51	0.16	3.91
Oreo	1.46	0.81	0.94	0.69	0.05	0.56
Peanuts	2.85	1.55	1.78	0.12	0.09	0.12
Forgone Sales of Stockouts	-98.60	-94.02	-101.82	-99.45	-95.39	-103.40
Increased Sales of Substitutes	35.99	9.08	32.90	29.81	2.88	25.81
Change in Total Sales	-62.61	-84.94	-68.92	-69.64	-92.50	-77.59
Percent Staying Inside	36.50	9.66	32.31	29.97	3.02	24.96
Change in Gross Profit	-31.43	-45.15	-36.66	-36.19	-47.36	-40.40

Table 7: Weekly Sales Impact of Simulated Stockout

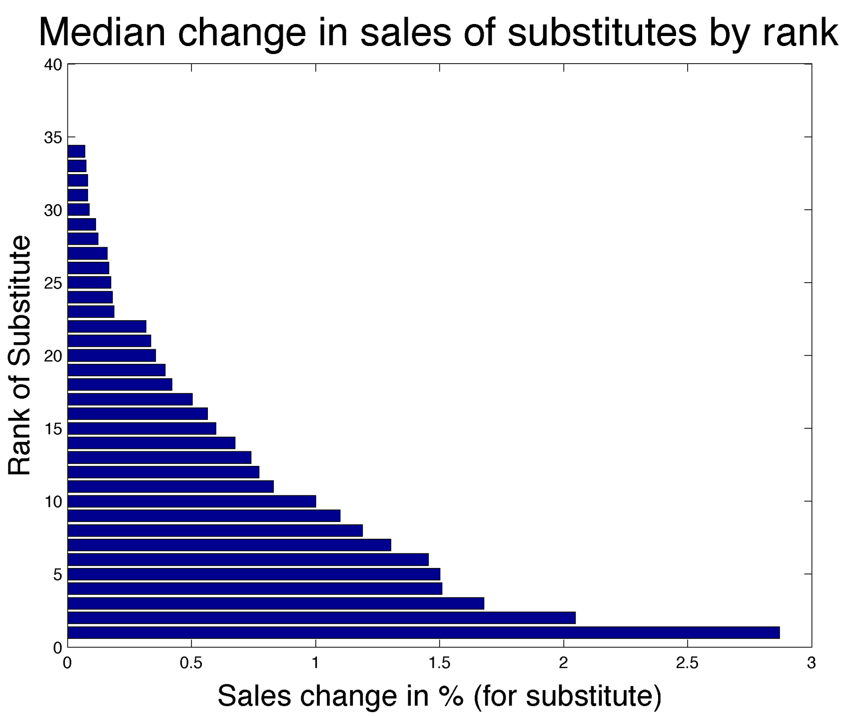

Figure 1: Median Change in Sales of Substitutes by Rank

A Multiple Stockouts, Alternative Methods, and Additional Results

A.1 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 is imputed). Conditional on the imputed values for the missing data, the M-step remains unchanged.

Let's suppose that we have K products which stockout in period t. We take the same approach as before and look to integrate out the α's, which tell us the fraction of consumers facing a particular availability regime. In the single stockout case there were two availability regimes which we parameterized as α and $1-\alpha$. Now if the are K unobserved stockouts, then there are 2^K possible availability regimes, and $2^K - 1$ parameters (since we require the weights sum to one).

$$E[q_{jti}] = \sum_{\forall \alpha_i : \sum \alpha = 1} y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)} g(\boldsymbol{\alpha}|\theta, y_{jt})$$

Where $\boldsymbol{\alpha} = [\alpha_0, \alpha_A, \alpha_B, \ldots, \alpha_{AB}, \ldots]$ is a vector of the appropriate 2^K values where subscripts denote stocked out products. Once again we can evaluate the summation 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, a week worth of data might be $M_t = 1000$ and $K = 5$ which 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 α vector implies a unique sequence of stockouts. If we had $K = 2$ products then $\boldsymbol{\alpha} = [\alpha_0, \alpha_A, \alpha_B, \alpha_{AB}]$, then 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 a α where $[0.3, 0.2, 0.2, 0.3]$, 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 a such a sequence is zero. It is helpful to define $\tilde{\boldsymbol{\alpha}} = f(\boldsymbol{\alpha})$ which represents the nonzero components of $\boldsymbol{\alpha}$ arranged in the order in which the stockouts occurred.

Likewise, we define $\boldsymbol{\omega} = [\omega_1, \omega_2, \ldots, \omega_K]$ as the beginning of period inventory for the K products which stocked out, where $\boldsymbol{\omega}$ is arranged in the order of the stockouts (just like $\tilde{\boldsymbol{\alpha}}$). It is also helpful to write $a(\alpha^{(k)}) = a_k$ which denotes the availability set corresponding to the k^{th} component of $\boldsymbol{\alpha}$. It is also helpful to define $\boldsymbol{p}(\theta, a(\alpha), x_t)$ 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.

What's nice is that we can define a finite recursive relationship for $g(\boldsymbol{\alpha}, \boldsymbol{\omega}, \theta, M_t)$, for the case where $K > 1$:

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

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 which 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 which did not stock out, thus there are only $K < J$ sums to evaluate.

A.2 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, statistical packages and the like. 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 trials m before ω successes of the first cell are observed. What makes it different from the negative binomial is that it also accounts for $q_2, \ldots, q_K$ successes of the next $K-1$ cells. The p.m.f. can be written:

$$NMult\left(m, \omega^{(1)}, q, \mathbf{p}(a,\theta)\right) = \frac{(m-1)!}{(m-1-\sum_k q_k)!\omega^{(1)} \dots q_K! q_0!} p_1^{\omega^{(1)}} p_2^{q_2} \dots p_k^{q_k} p_0^{m - \sum_{i=0}^{K} q_i}$$

A.3 Alternative Computational Methods

In the case where integrating over the exact distribution $g(\boldsymbol{\alpha}|\cdot)$ at all values of the domain is prohibitive, there are some alternatives. Since we're using the distribution to compute an expectation, we can compute the E-step numerically without a problem. Recall that the likelihood is linear in the sufficient statistics, so any approach to numeric integration where the approximation error is mean zero should be acceptable. Also recall the form of the E-step.

$$E[q_{jti}] = \sum_{\forall \alpha_i : \sum \alpha = 1} y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)} g(\boldsymbol{\alpha}|\theta, y_{jt})$$

This can be easily approximated by linear functions since it is of the form $\frac{f(z_i)}{\sum_i f(z_i)}$, and 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 small number of points at which the function is evaluated.

A Monte Carlo approach could involve sampling from $g(\cdot)$ in a random or quasi random way, evaluating the density at those points, and then evaluating the integral numerically at a finite number of points. Such an approach should yield consistent estimates for the expected sufficient statistics so long as the draws are held fixed across products within a stockout interval. 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. This requires that we know the probability of different stockout patterns (A then B then C, etc.) in order to weight our draws appropriately. When computing the probability of different stockout orders becomes difficult, continuous approximations generally become more reasonable. Most of these results are well described in the literature on queuing theory and operations research.

For extremely high dimensional problems we might find that even drawing from $g(\cdot)$ becomes too burdensome. In this case we could consider simulating consumer purchases (since we know the sales are distributed as a multinomial for a given set of parameters θ). It is important that we only track those products stocked out and an "all other goods" option. In this case simulate consumers until all of the products that stocked out in the data have stocked out during that period, and we can count the fraction of all other good consumers facing each availability set and compute α directly. The key trick is that we can use the "all other goods" option rather than having to keep track of other sales for each product, this means we do not have to use an accept-reject method for simulation, but rather can keep every draw. It also means that we can expect consistent results even for a finite number of simulated chains of consumers. 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. [47]

A.4 Latent Types

As discussed in section 3.5, one might worry that our method relies too heavily on the assumption of exchangeability 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 then 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.[48] 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).

$$\begin{aligned}
p_{jtA}(x, \delta, \mu_A, \sigma_A) &= \int \frac{\exp(\delta_j - \alpha_i x_{jt})}{1 + \sum_{k \in a_t} \exp(\delta_k - \alpha_i x_{kt})} \phi(\alpha_i | \mu_A, \sigma_A) \\
p_{jtB}(x, \delta, \mu_B, \sigma_B) &= \int \frac{\exp(\delta_j - \alpha_i x_{jt})}{1 + \sum_k \exp(\delta_k - \alpha_i x_{kt})} \phi(\alpha_i | \mu_B, \sigma_B)
\end{aligned}$$

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

[47] The reason the "all other goods" formulation is acceptable follows from the so-called "Law of Small Numbers" which describes how locally all sales behave as independent poissons over short intervals when the p_j's are fixed.

[48] In 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.

$$p_{jt}(x,\theta) = \gamma \cdot s_{jtA}(x,\delta,\mu_A,\sigma_A) + (1-\gamma) \cdot s_{jB}(x,\delta,\mu_B,\sigma_B)$$

We could estimate this model using a variety of techniques, including ML, where $\theta = [\delta, \mu_A, \mu_B, \sigma_A, \sigma_B, \gamma]$. If s_{jtX} is computed via simulation, then we call the procedure MSL: [49]

$$\hat{\theta} = \arg\max_{\theta} \sum_{jt} y_{jt} \ln p_{jt}(x,\theta)$$

This example highlights the fact that likelihood-based approaches are always available for describing multinomial data.[50] The challenge with any approach (likelihood or otherwise) will be in representing the choice probabilities correctly. Furthermore, we require many draws when p_{jt} is simulated, because the log function is nonlinear. We usually require that $\frac{ns}{n} \to \infty$ and note that MSL is inconsistent for a fixed number of draws.

Another simple estimator would be a method of moments (or method of simulated moments) approach, in which we utilize the discrepancy between the predicted and actual sales, and impose the condition that the prediction error is orthogonal to our set of instruments. Fortunately, the prediction error is a linear function of $p_{jt}(\theta)$:

$$\hat{\theta}_{MSM} = \arg\min_{\theta} (q_{jt} - M_t \cdot p_{jt}(\theta)) * z_{jt}.$$

The optimal instruments minimize the variance of the MSM estimator, and are given by the score, or $z_{jt} = \frac{\partial \ln p_{jt}(\theta)}{\partial \theta}$. When these instruments are used $\hat{\theta}_{MSM} = \hat{\theta}_{ML}$.(Train forthcoming) Simulating the score directly involves a $\frac{1}{p_{jt}}$ term and requires using an Accept-Reject simulator. A simpler (but less efficient) estimator would exploit the fact that prediction error should be uncorrelated with the set of available products and their characteristics. In order to pin down σ parameters, one could also include higher moments of characteristics.

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.[51]

[49] It turns out optimization for these problems is actually quite difficult, even though writing them down is easy.

[50] The multinomial distribution is often considered a "non-parametric" distribution for this reason.

[51] Relatedly, 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

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 $p_j(a, \theta_s), p_j(a, \theta_a)$ respectively, and after the stockout as $p_j(b, \theta_s), p_j(b, \theta_a)$. 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; λ 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:

$$\frac{q_{jt}}{M_t} = \lambda \left[\gamma_a p_j(a, \theta_s) + (1 - \gamma_a) p_j(a, \theta_a)\right] + (1 - \lambda) \left[\gamma_b p_j(b, \theta_s) + (1 - \gamma_b) p_j(b, \theta_a)\right].$$

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 λ (conditional negative binomial), which is easy to integrate out. We could use a similar technique if we fixed:

$$\begin{aligned}
\tilde{p}_j(a, \theta) &= \gamma_a p_j(a, \theta_s) + (1 - \gamma_a) p_j(a, \theta_a) \\
\tilde{p}_j(b, \theta) &= \gamma_b p_j(b, \theta_s) + (1 - \gamma_b) p_j(b, \theta_a).
\end{aligned}$$

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

The alternative approach implies imputing sales for all four cases. However, the resulting integrals are over the joint distribution of λ, γ. While we knew the marginal distribution of λ, 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

data relevant to the choice probability computation is unobserved, except now it is the a_t's. 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).

times for each stockout scenario would dramatically increase the parameter space. We would probably want to enforce some additional conditions on γ in this approach. 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 (λ, γ).

A.5 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 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 $a'_t = \{a_t\} \setminus \{l\}$.

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

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.[52] 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.

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

[52] In 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.

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

A.6 Comparison of Likelihoods Under Alternative Choices for $g(\cdot)$

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. Recall equation (9):

$$L(\mathbf{y}|\mathbf{a},\mathbf{x},\theta) = l(\mathbf{y_{obs}}|\mathbf{a_{obs}},\mathbf{x_{obs}},\theta) + \int l(\mathbf{y_{mis}}|\mathbf{a_{mis}},\mathbf{x_{mis}},\theta) g(\mathbf{a_{mis}}|\mathbf{a_{obs}},\mathbf{y},\mathbf{x},\theta) \partial \mathbf{a_{mis}}$$

This is the true likelihood of the observed data $(\mathbf{a},\mathbf{x},\mathbf{y})$ given the parameter θ. Therefore any θ that maximizes $L(\cdot)$ yields a consistent estimate of the true θ_0.

The integral in (9) is not easy to solve. Standard approaches do not solve (9), but rather solve some other problem by substituting in a different $g(\cdot)$. For example, in the case where we ignore the missing data it would be as if we set $g(\cdot) = 0$ everywhere (which is not a proper density distribution). Or in the case of full availability it would be as if we set $g(\cdot)$ to be a delta function that took on value 1 only at the full availability value of a. Assuming stockouts happen at the beginning or the end of the period places similar structure on $g(\cdot)$ (making it a delta function). The problem with this is that $g(\cdot)$ does not have any free parameters, but is completely specified by the demand parameters as a conditional negative binomial.

For example, let λ denote the fraction of consumers arriving before the stockout and $1-\lambda$ denote the fraction of consumers arriving after the stockout. We could compute the expected sales for each product before and after the stockout:

$$\begin{aligned} E[q_j^{before}] &= q_j \int \frac{\lambda p_j}{\lambda p_j + (1-\lambda)\tilde{p}_j} f(\lambda) \partial \lambda \;\forall j \\ E[q_j^{after}] &= q_j - E[q_j^{before}] \end{aligned}$$

This is essentially what we do in the E-Step of our EM procedure. An alternative might be to consider the marginal data augmentation framework of Tanner and Wong (1987), in which we think of the stockout time λ 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 $\hat{\lambda}$ such that:

$$q_j \frac{\hat{\lambda} p_j}{\hat{\lambda} p_j + (1-\hat{\lambda}) p_j} = q_j \int \frac{\lambda p_i}{\lambda p_j + (1-\lambda)\tilde{p}_j} f(\lambda) \partial \lambda$$

If this were true we could treat $\hat{\lambda}$ as an additional parameter to estimate. Unfortunately, we don't have a single equation, but rather a set of J equations, and only a single λ. Thus only in very special (degenerate) cases can a single $\hat{\lambda}$ satisfy all J equations. [53] That is, any $\hat{\lambda}$ which gives a consistent expectation of some j will not give a consistent expectation for some other j'.[54]

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.

A.7 Additional Results

Table 8 reports the estimates of the product dummies from each of the models estimated on the base dataset. Table 9 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. (However, the likelihood from this model is lower than that under the nested-logit specification.)

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

[54] Notice that this is true for a generic $f(\theta)$. As we've noted previously, stockouts imply a particular distribution on $f(\cdot)$, which also prevents λ from being a free parameter.

	Single Category Nested			Five Category Nested			Random Coefficients		
	Full	Ignore	EM	Full	Ignore	EM	Full	Ignore	EM
PopTart	-5.49	-6.29	-5.77	-6.93	-6.36	-5.97	-12.70	-6.18	-11.35
Choc Donuts	-5.52	-6.18	-5.69	-6.98	-6.25	-5.87	-8.95	-6.08	-8.11
Ding Dong	-5.58	-6.39	-5.83	-7.18	-6.47	-6.04	-11.12	-6.28	-9.96
Banana Nut Muffin	-5.65	-6.62	-6.00	-7.40	-6.71	-6.24	-14.20	-6.49	-12.62
Rice Krispies	-5.80	-7.01	-6.28	-7.85	-7.12	-6.57	-7.76	-6.86	-7.51
Pastry	-5.58	-6.45	-5.88	-7.25	-6.54	-6.10	-10.92	-6.33	-9.85
Gma Oatmeal Raisin	-5.75	-6.62	-6.08	-5.26	-6.07	-5.73	-11.61	-6.52	-10.56
Chips Ahoy	-5.83	-6.78	-6.19	-5.28	-6.17	-5.81	-8.37	-6.67	-7.94
Nutter Butter Bites	-5.99	-7.13	-6.44	-5.33	-6.38	-5.98	-8.07	-6.99	-7.78
Knotts Raspberry Cookie	-6.00	-7.14	-6.45	-5.34	-6.39	-5.99	-9.21	-7.00	-8.68
Gma Choc Chip	-5.65	-6.54	-5.98	-5.21	-5.96	-5.64	-8.88	-6.42	-8.30
Gma Mini Cookie	-5.84	-6.70	-6.17	-5.29	-6.17	-5.82	-7.55	-6.60	-7.28
Gma Caramel Choc Chip	-5.81	-6.54	-6.07	-5.28	-6.06	-5.75	-10.73	-6.45	-9.75
Rold Gold	-5.12	-6.25	-5.54	-4.40	-7.00	-5.65	-9.47	-6.10	-8.73
Sunchip Harvest	-5.15	-6.33	-5.59	-4.41	-7.11	-5.71	-6.67	-6.18	-6.48
Dorito Nacho	-5.22	-6.53	-5.74	-4.43	-7.38	-5.87	-6.96	-6.36	-6.79
Cheeto Crunchy	-5.22	-6.50	-5.72	-4.43	-7.35	-5.85	-8.04	-6.34	-7.63
Gardetto Snackens	-4.96	-6.03	-5.37	-4.35	-6.71	-5.47	-7.80	-5.90	-7.38
Ruffles Cheddar	-5.31	-6.71	-5.87	-4.46	-7.63	-6.01	-7.59	-6.54	-7.32
Fritos	-5.53	-7.18	-6.19	-4.53	-8.26	-6.35	-7.50	-6.97	-7.33
Lays Potato Chip	-5.58	-7.25	-6.26	-4.55	-8.36	-6.43	-7.67	-7.04	-7.47
Munchies Hot	-5.24	-6.55	-5.77	-4.44	-7.38	-5.90	-7.15	-6.38	-6.95
Misc Chips 2	-5.68	-7.45	-6.39	-4.58	-8.63	-6.57	-7.62	-7.23	-7.47
Munchies	-5.45	-6.99	-6.05	-4.51	-8.02	-6.21	-8.03	-6.79	-7.71
Misc Chips 1	-5.44	-6.93	-6.03	-4.51	-7.90	-6.17	-7.37	-6.74	-7.15
Dorito Guacamole	-5.51	-7.16	-6.18	-4.53	-8.25	-6.34	-7.20	-6.95	-7.11
Snickers	-4.75	-5.53	-5.05	-5.21	-5.15	-4.69	-9.01	-5.44	-8.22
Twix	-4.90	-5.87	-5.28	-5.47	-5.39	-4.84	-9.08	-5.75	-8.33
M&M Peanut	-5.05	-6.15	-5.48	-5.72	-5.58	-4.97	-8.91	-6.01	-8.23
Reeses Cup	-5.41	-6.90	-6.02	-6.33	-6.12	-5.31	-8.85	-6.71	-8.33
Kit Kat	-5.46	-7.00	-6.09	-6.41	-6.19	-5.35	-10.83	-6.80	-9.94
Caramel Crunch	-5.46	-6.99	-6.09	-6.41	-6.19	-5.36	-9.27	-6.80	-8.69
M&M	-5.40	-6.89	-6.01	-6.33	-6.11	-5.30	-10.95	-6.70	-10.02
Hershey Almond	-5.60	-7.28	-6.29	-6.65	-6.39	-5.49	-8.78	-7.06	-8.35
Babyruth	-5.63	-7.27	-6.30	-6.68	-6.41	-5.50	-11.93	-7.07	-10.88
Starburst	-5.53	-6.34	-5.83	-5.10	-5.59	-5.45	-12.60	-6.25	-11.30
Kar Nut Sweet/Salt	-5.69	-6.64	-6.05	-5.15	-5.74	-5.59	-8.42	-6.52	-7.95
Snackwell	-5.96	-7.20	-6.45	-5.25	-6.00	-5.83	-8.82	-7.04	-8.38
Skittles	-5.58	-6.50	-5.92	-5.11	-5.64	-5.49	-15.04	-6.38	-13.35
Payday	-5.84	-6.89	-6.26	-5.21	-5.89	-5.73	-8.97	-6.77	-8.45
Oreo	-6.22	-7.73	-6.83	-5.34	-6.24	-6.06	-10.38	-7.54	-9.74
Peanuts	-5.89	-7.13	-6.38	-5.22	-5.92	-5.76	-6.98	-6.97	-6.94
Peter Pan (Crck)	-5.99	-7.20	-6.48	-5.27	-6.04	-5.86	-8.10	-7.06	-7.85
Hot Tamales	-5.81	-6.67	-6.12	-5.20	-5.81	-5.65	-12.08	-6.57	-10.82

Table 8: d_j Parameters

	Single Parameter Nested			Category Specific Nested			Random Coefficients		
	Full	Ignore	EM	Full	Ignore	EM	Full	Ignore	EM
Constant	-5.62	-7.25	-6.30	-4.63	-6.81	-5.74	-6.30	-7.04	-6.41
	(0.19)	(0.27)	(0.21)	(0.46)	(0.37)	(0.21)	(0.37)	(0.24)	(0.32)
Calories	5.10	6.55	5.76	6.75	-3.04	2.17	7.84	6.35	7.71
	(2.21)	(3.05)	(2.44)	(5.23)	(4.26)	(2.36)	(4.25)	(2.70)	(3.72)
Fat	-2.59	-3.37	-2.93	-4.71	0.37	-1.74	-3.01	-3.26	-3.06
	(1.09)	(1.50)	(1.20)	(2.57)	(2.10)	(1.16)	(2.09)	(1.33)	(1.83)
Sodium	0.09	0.07	0.10	-1.14	0.38	0.08	-3.73	0.08	-3.01
	(0.41)	(0.56)	(0.45)	(0.97)	(0.79)	(0.44)	(0.79)	(0.50)	(0.69)
Carbs	-2.47	-2.33	-2.40	-0.76	0.54	-1.57	-1.91	-2.35	-2.00
	(1.23)	(1.70)	(1.36)	(2.92)	(2.38)	(1.31)	(2.37)	(1.51)	(2.07)
Sugar	-0.14	-0.01	-0.09	-3.30	2.29	0.66	-8.88	-0.02	-7.17
	(0.50)	(0.69)	(0.55)	(1.18)	(0.96)	(0.53)	(0.96)	(0.61)	(0.84)
Choc	0.17	0.21	0.19	0.14	0.44	0.41	0.84	0.20	0.75
	(0.14)	(0.20)	(0.16)	(0.33)	(0.27)	(0.15)	(0.27)	(0.17)	(0.24)
Cheese	0.18	0.17	0.18	0.41	0.01	0.18	-0.19	0.17	-0.12
	(0.16)	(0.22)	(0.18)	(0.38)	(0.31)	(0.17)	(0.31)	(0.20)	(0.27)
R^2	0.218	0.253	0.219	0.525	0.608	0.480	0.917	0.244	0.900

Table 9: d_j's on Characteristics

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 Eocnomic 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 Compettitve 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.

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 Appraoch to Estimating Market Equilibrium Models," *Management Science*, 50(5), 605–616.

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63(4), 891–951.

Hadley, G., and T. Whitman (1963): *Analysis of Inventory Systems*. Prentice Hall.

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. Lundsqvist, 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.

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).

Train, K. (forthcoming): "EM Algorithms for Nonparametric Estimation of Mixing Distributions," *Journal of Choice Modeling.*

www.vending.org (2008): "Vending 101," www.vending.org/industry/vending101.pdf.

Zeithammer, R., and P. Lenk (2006): "Less data sometimes contains more information: the relationship between choice-subsets and precision of inference," Working Paper.