IO
7/21/07
1:15 pm

Demand Estimation Under Incomplete Product Availability*

Christopher T. Conlon
Julie Holland Mortimer

July 12, 2007

Abstract

Stockout events are a common feature of retail markets. When stockouts change the set of available products, observed sales provide a biased estimate of demand. If a product sells out, actual demand may be greater than observed sales, leading to a negative bias in demand estimates. At the same time, sales of substitute products may increase. Such events generate variation in choice sets, which is an important source of identification in the IO literature. In this paper, we develop a simple procedure that allows for variation in choice sets within a "market" over time using panel data. This allows for consistent estimation of demand even when stockouts imply that the set of available options varies endogenously. We estimate demand in the presence of stockouts using data from vending machines, which track sales and product availability. When the corrected estimates are compared with naive estimates, the size of the bias due to ignoring stockouts is shown to be large.

*We thank Susan Athey, Steve Berry, Uli Doraszelski, Phil Haile, Ken Hendricks, Phillip Leslie, and Ariel Pakes for helpful discussions and comments. Financial support for this research was generously provided through NSF grant SES-0617896.

1 Introduction

Retail and service markets account for 30% of GDP and 48% of employment in the US, yet most economic models assume that retail settings are unimportant for understanding consumer demand and firms' decisions. Specifically, most methods of demand analysis rely on the assumption that all products are available to all consumers. While many industries, such as automobiles and computer chips, have been successfully analyzed under this assumption (Berry, Levinsohn, and Pakes 1995), research into many retail markets suggests that retail settings are characterized by important deviations from this model. Specifically, "stock-outs" of products, or periods where products are unavailable are common in many settings. Furthermore, both producers and consumers identify product availability as an important consideration in these markets. When the goods in question are perishable or seasonal, or generally lack inter-temporal substitutability, management of inventory is not an ancillary concern; it is the primary problem that firms address. When stockouts change the set of available products, observed sales provide a biased estimate of demand for two reasons. The first source of bias is the *censoring* of demand estimates. If a product sells out, the actual demand for a product 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 the true demand for these goods.

The current class of discrete choice models prevalent in the IO literature is able to address variation in the choice sets facing consumers across markets. In fact, variation in choice sets across markets is an important source of identification in these models. In this paper, we develop a simple procedure that allows for variation in choice sets within a "market" over time using panel data. This allows for consistent estimation of demand even when stockouts imply that the set of available options varies endogenously.

If the choice set facing the consumer were observed when each choice was made, correcting demand estimates would be simple. However, in many real world applications inventories are only observed periodically. This presents an additional challenge for estimation, because the regime under which choices took place must be estimated in addition to parameters. Thankfully, this is a well understood missing data problem and the EM algorithm of Dempster, Laird, and Rubin (1977) applies.

The dataset that we use tracks the sales of snack foods in vending machines located on the campus of Arizona State University (ASU). Wireless observations of the sales and product availability, along with numerous and repeated observations of stock-outs, make this dataset well-adapted to the analysis of product availability.

When the corrected estimates are compared with the naive estimates, the size of the bias is shown to be large, and the welfare implications of stockouts would be substantially mismeasured with naive estimates. This paper focuses only on the static analysis of *demand* in the presence of reduced product availability. It does not consider either dynamic interactions or the problem of the retailer.

2 Relationship to Literature

The differentiated products literature in 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 Musalem, Bradlow, and Raju (2006). While this paper uses a common Bayesian technique to address missing data, it is not a fully Bayesian model.

There is also a substantial literature in IO on the dynamics of price and inventory. Previous studies have looked at the effect of coupons and sales on future demand in terms of "consumer inventories" (durable goods) (Nevo and Hendel 2007b, Nevo and Hendel 2007a, Nevo and Wolfram 2002). And other studies have looked at the dynamic interaction between retailer inventories (and the cost of holding them) and the markups extracted by the retailer (Aguirregabiria 1999). While retailer inventories are explicitly modeled in the example we examine, these sorts of dynamics are not an issue because the retailer does not have the ability to dynamically alter the control (price or product mix). In fact, vending is a useful industry to study product availability precisely because we need not worry about these other dynamic effects.

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 E-M techniques (Dempster, Laird, and Rubin 1977) to estimate the Poisson model in the presence of missing choice set data.

This paper aims to connect these two 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.

2.1 Inventory Systems

When talking about inventory systems we use the standard dichotomy established by Hadley and Whitman (1963). The first type of inventory system 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. This system is also known as "real-time" inventory.[1]

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 real-time inventory systems are becoming more common in retailing environments thanks to innovations in information technology, most retailers still do not have access to real-time inventory data. However, periodic inventory systems can be used to approximate perpetual data. As the size of the sampling period becomes sufficiently small, periodic data approaches perpetual data. In the limit where the inventory is sampled between each transaction, this is equivalent to having real-time data. These two points become very important in the estimation section.

Sampling inventory more frequently helps to mitigate limitations of the periodic inventory system. However, the methodological goal of this paper is to provide consistent estimates of demand not only for perpetual inventory systems but for periodic ones as well.

3 Model

A typical starting point in discrete choice models is to begin by writing down a consumer's utility function. However, instead of doing that we consider what the observed data look like, and try to write down an expression for the likelihood. Assume that the observed data are broken up into time periods $t = 1, 2, \ldots, T$ and that products are denoted by the subscript $j = 1, 2, \ldots, J$. For each product, denote y_{jt} as the quantity sold of the product j in period t. It is important to note that the precise order of sales may be unobserved, with only aggregate data available for each period t.

We assume that in each period t, and there are M_t potential consumers.[2] This is a typical assumption in the differentiated products literature (M_t is often based on census data such as in Berry, Levinsohn, and Pakes (1995) or Nevo (2001)). In this section we assume that the set of available products a_t is constant within a period. We relax this assumption in the next section.

[1] 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).

[2] This assumption can be relaxed later.

Assumption 1. *(Discrete Choice) Each of the M_t consumers in period t must either choose some product $j \in a_t$ or the outside good $j = 0$.*

Assumption 2. *(Independence of Consumers) Each consumer's preferences are independent of other consumer's preferences (and choices) and each set of preferences is the realization of an i.i.d. draw from some stable population distribution (perhaps conditional on some x_t)*

Assumption 1 states that each of the M_t consumers face a choice: they may buy exactly one product j in the set a_t of products available in their market, or they may choose to buy nothing at all. If M_t is known, then so is the number of consumers who did not buy a product. The outside good is denoted $y_{0t} = M_t - \sum_{j \in a_t} y_{jt}$.

Assumption 2 is not a new assumption to this literature either. It implies that consumer preferences are independent of one another. This may seem contrary to the nature of stockouts but it isn't. Stockouts highlight the important distinction between the primitives, the underlying (latent) preferences of consumers, and the observable purchase decisions. Assumption 2 requires that preferences are i.i.d. draws from the distribution of preferences, while the observed decisions are realizations not only of preferences, but also of choice sets, which can be dependent on the preferences (and decisions) of other consumers. In analyses where stockouts are not directly addressed these are often conflated.

Assumption 2 also means that the population of consumer preferences we're sampling from cannot change within a period of observation. If such a change were to happen we can only make inferences about the overall mixture, not its components. For example, if we observed data on sales from 4-8pm, and at 5pm the population of consumers changes, then 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. This sort of heterogeneity can be addressed in our approach (as part of the observable x_t), but not within a single period of observation. This does not mean that there is no room for unobservable heterogeneity–consumer preferences can be explained by a distribution–but conditional on an (θ, x_t), that distribution must be the same.[3]

For simplicity, let $\mathbf{y_t} = [y_{0t}, y_{1t}, y_{2t}, \ldots, y_{Jt}]$. Then for each market, the data provide information on $(\mathbf{y_t}, M_t, a_t, x_t)$ where x_t is some set of exogenous explanatory variables. 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

[3]This is not a new limitation for the discrete choice literature, but it is more salient when we try to use the discrete choice approach for obtaining high frequency estimates of consumer demand. Previous studies have relied on annual or quarterly data. In these sorts of datasets it is pretty clear that for each observation, short-term heterogeneity in the population gets "averaged out" in the overall distribution of consumer preferences.

is that we can reorder the unobserved purchase decisions of individual consumers within a period t. Now, we apply assumption 2 again and the fact that M_t is known to write the likelihood function as a multinomial with parameters $n = M_t$, and $p = [p_{1t}, p_{2t}, \ldots]$

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

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 $C(M_t, \mathbf{y_t})$ 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.

Assumption 3. *(Independence of Periods/Locations) Each period t is independent of other periods, such that for a given θ, $p_j(\theta, a_t, x_t)$ is the same function across t and depends only on a_t the set of available products (as well as the exogenous variables x_t).*

One way to look at this is an identification condition on $p_j(\theta, a_t, x_t)$, that it is fixed and completely characterized by its parameters.[4] This assumption now lets us consider the joint likelihood of several periods as the product of $f(\mathbf{y_t}|\theta, M_t)$ over all periods $t = 1, 2, \ldots, T$. In the shorthand notation below, boldface denotes vectors over all periods t. Thus we define $\mathbf{y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_T}]$, $\mathbf{M} = [M_1, \ldots, M_T]$, $\mathbf{a} = [a_1, \ldots, a_T]$, and $\mathbf{x} = [x_1, \ldots, x_T]$.

$$\begin{aligned} L(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &= \prod_{\forall t} C(M_t, \mathbf{y_t}) p_{0t}^{y_{0t}} p_{1t}^{y_{1t}} \ldots p_{Jt}^{y_{Jt}} \\ &\propto \prod_{\forall t} \prod_{\forall j \in a_t} p_{jt}^{y_{jt}}(\theta, a_t, x_t) \\ l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \sum_{\forall j \in a_t} y_{jt} \ln p_j(\theta, a_t, x_t) \end{aligned} \tag{3}$$

Given a function $p(\cdot)$ known up to a set of parameters θ we can now consider estimating this model by maximum-likelihood type procedures. With this in mind we provide a straightforward result for sufficient statistics required in estimation.

Theorem 1. *If p_{jt} is a deterministic function of the set of available products and the unknown parameters (and some observable exogenous x_t), and a_t is constant across a period t, then for some $a_t = a$, $\mathbf{q_{a,x}} = \sum_{t: a_t = a, x_t = x} \mathbf{y_t}$ is a sufficient statistic for the contribution of $T_{a,x} = \{t : a_t = a, x_t = x\}$'s contribution to the likelihood.*

[4] A formal identification condition requires that the set of $p_j(\theta, a_t, x_t)$'s are uniquely generated by a θ.

Proof:

Since (a_t, x_t) are fixed within the period t we have:

$$\begin{aligned}
l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \sum_{\forall j \in a_t} y_{jt} \ln p_j(\theta, a_t, x_t) \\
&= \sum_{\forall (a,x)} \sum_{\forall t:(a_t,x_t)=(a,x)} \sum_{\forall j \in a_t} y_{jt} \ln p_j(\theta, a, x) \\
&= \sum_{\forall (a,x)} \sum_{\forall j \in a} \ln p_j(\theta, a, x) \sum_{\forall t:(a_t,x_t)=(a,x)} y_{jt} \\
&= \sum_{\forall (a,x)} \sum_{\forall j \in a} q_{j,a,x} \ln p_j(\theta, a, x) \qquad (4)
\end{aligned}$$

Corollary to Theorem 1. *Since the likelihood is additively separable in the sufficient statistics $\mathbf{q_a}$, the sums $\mathbf{q_a}$ can be broken up in an arbitrary way, including one sale at a time, as it will not affect the likelihood so long as the sales are of the same (a, x) regime. It is also clear that within an (a, x) regime the order of sales does not affect the sufficient statistic $q_{a,x}$*

4 Adjusting for Stockouts

4.1 Perpetual Inventory

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_t, a'_t). This case is illustrated in figure 1.

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). We can see this by introducing the subscript l which denotes a single consumer. Though we might not observe the order in which the consumers make purchases, perpetual inventory implies that (y_l, a_l, x_l) is known.[5] If we return to our likelihood equation we see that it remains unchanged when we consider

[5]Technically we also have that $x_l = x_t$, that the observables don't vary across individuals, which can be relaxed with micro data, and perhaps otherwise, but is beyond the scope of this paper.

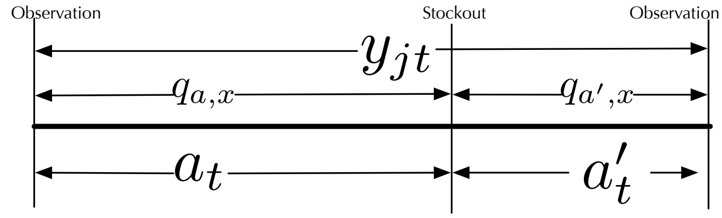

Figure 1: For perpetual inventory, sufficient statistics are directly observed.

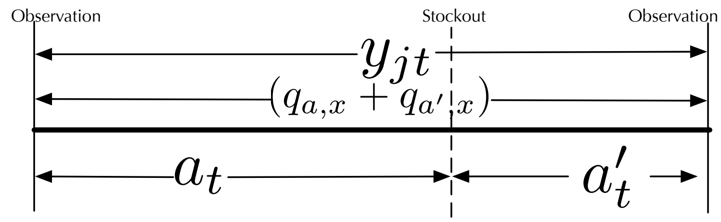

Figure 2: For periodic inventory, sufficient statistics are unobserved.

single consumers instead of time periods (Corollary 1).

$$
\begin{aligned}
l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall t} \left(\sum_{\forall j \in a_t} \ln p_j(\theta, a_t, x_t) \sum_{\forall l:(a_l,x_l)=(a_t,x_t)} y_{jl} + \sum_{\forall j \in a'_t} \ln p_j(\theta, a'_t, x) \sum_{\forall l:(a_l,x_l)=(a'_t,x)} y_{jl} \right) \\
l(\mathbf{y}|\theta, \mathbf{M}, \mathbf{a}, \mathbf{x}) &\propto \sum_{\forall (a,x)} \sum_{\forall j \in a} \ln p_j(\theta, a, x) \sum_{\forall l:(a_l,x_l)=(a,x)} y_{jl} \\
&= \sum_{\forall (a,x)} \sum_{\forall j \in a} q_{j,(a,x)} \ln p_j(\theta, a, x) \qquad (5)
\end{aligned}
$$

The sufficient statistic representation makes it clear that all we need to observe are the aggregate sales of each product $q_{j,a,x}$ for a given (a, x) regime. In other words, so long as the sales are attributed to the correct choice set, we can hope to estimate $p(\cdot)$.

4.2 Periodic Inventory

As has already been discussed, many retail environments observe inventories periodically. This presents additional challenges when investigating stockouts, because thus far we've relied on the fact that the choice set a was observed when each sale was made. Now if a stockout takes place in period t, the availability is known only at the beginning and the end of the period. Like in the perpetual case, we could denote the set of available choices at the beginning of period t by a_t, and the set remaining at the end of t by a'_t. What we'd like to do is assign the sales in period t to each (a_t, a'_t) regime, but because inventory is observed only periodically, we can't. As illustrated in figure 2, we do not observe which sales occurred during a_t and which during a'_t. Instead we observe only the sum, $\mathbf{y_{jt}} = \mathbf{q_{a_t,x_t}} + \mathbf{q_{a'_t,x_t}}$.

A natural division for the dataset is to label each period t as a period of fully observed availability, $t \in T_{obs}$ or a period where availability is only partially observed, $t \in T_{mis}$. For periods where availability is fully observed (complete data), a_t is known and is constant across the period t. For periods where availability is partially observed (incomplete data), we only know that a_t at the beginning and the end of the period, and that it takes on values in some sequence of availability sets $[a_s, a'_s]$.[6] More formally, we can break up the dataset along this partition:

$$\begin{aligned} T_{obs} &= \{t : a_t = a_s\} \\ T_{mis} &= \{t : a_t = [a_s, a'_s]\} \end{aligned}$$

We can break up the observed sales $\mathbf{y_t}$ along the same partition where $\mathbf{y_{obs}} = \{\mathbf{y_t} : t \in T_{obs}\}$ and $\mathbf{y_{mis}} = \{\mathbf{y_t} : t \in T_{mis}\}$. We recall that (3) gave us the nice property that the log-likelihood function was linear in periods, t. This lets us rearrange terms so that we can partition the log-likelihood into the complete and incomplete data contributions:

$$\begin{aligned} l(\mathbf{y}|\theta, \mathbf{M}) \propto & \quad \textstyle\sum_{\forall t} \sum_{\forall j \in J_t} y_{jt} \ln p_{jt}(\cdot) & (6) \\ = & \quad \underbrace{\sum_{\forall t \in T_{obs}} \sum_{\forall j \in J_t} y_{jt,obs} \ln p_j(\theta, a_t, x_t)} + \underbrace{\sum_{\forall t \in T_{mis}} \sum_{\forall j \in J_t} y_{jt,mis} \ln p_{jt}} & (7) \\ = & \quad l_{obs}(\mathbf{y_{obs}}|\theta, \mathbf{M}) + l_{mis}(\mathbf{y_{mis}}|\theta, \mathbf{M}) & (8) \end{aligned}$$

It's easy to see that the first term $l_{obs}(\cdot)$ is just the complete data contribution to the log-likelihood, and it doesn't differ at all from the case with perpetual inventory or no stockouts. In fact, computing $q_{a,x}$ for that segment of the dataset is straightforward, and proceeds just like the perpetual inventory case. The second term $l_{mis}(\cdot)$ is more difficult to deal with, because we do not observe the set of available products and the choice probabilities from (1) can no longer be written as $p_{jt} = p_j(\theta, a_t, x_t)$. Another way to look at this is to see that in this scenario that the sufficient statistic $\mathbf{q_{a,x}}$ is no longer known, since sales occur that cannot be matched to a single a regime. This is what's known as a *missing data problem*.

The question becomes, "Precisely which data are missing?" One might be inclined to treat the sales $\mathbf{y_t}$ as observed and the availability regimes (and $p_j(\theta, a_t, x_t)$) as unobserved, however, the correct approach is to consider the sufficient statistics and realize that while $\mathbf{y_t}, a_t, a'_t$ are observed $\mathbf{q_{mis,a,x}}$ are not. In other words, the sales under each availability regime are missing (though we know their sum). In fact, in most discrete choice problems, we can think about individual purchases as being missing data.[7] We don't usually worry about the missing data problem in that context, because while the individual purchases are not directly observable from aggregate data, the sufficient statistics are.[8] However, in the case of stockouts with periodic inventory, not only are individual purchases not observed,

[6] It is clear that perpetual inventory represents the special case where all $t \in T_{obs}$ and $T_{mis} = \emptyset$.

[7] This issue is addressed in the case of aggregate data (but without stockouts) by Musalem, Bradlow, and Raju (2006).

[8] An important exception is when we have micro data, then sufficient statistics vary with consumer observables x_t, and the identities of consumers matter.

but neither are the sufficient statistics.

Incorporating Missing Data

One way to deal with a missing data problem is to to ignore $l_{mis}(\cdot)$ and just do estimation on the $l_{obs}(\cdot)$ part. This is akin to throwing away the observations $\mathbf{y_{mis}}$ and just doing estimation when the set of available products is fully observed. There's a large literature on when this strategy leads to consistent estimates (Rubin 1976) . In the multinomial case, so long as discarding the data doesn't affect the distribution of the sufficient statistics, ignoring the missing data should only lead to a loss of efficiency, not inconsistency.

However, it is easy to see that the sufficient statistics are affected by discarding the missing data, with the following example: Consider the product that stocks out. If it has capacity ω, then if the data are observed we know that $q < \omega$, and if the data are discarded we know that $q = \omega$. Thus the sufficient statistic clearly depends on the "missingness" of the data. Therefore, by ignoring the missing data, we would expect to systematically underestimate demand for products which stock out.

Another approach is to replace $l_{mis}(\cdot)$ with some consistent estimator that is computable. One can replace $l_{mis}(\cdot)$ with its expectation by integrating it out over all possible values of the missing data. The typical notation here is:

$$Q(\theta) \quad = \quad l_{obs}(\mathbf{y_{obs}}, \theta) + E[l_{mis}(\mathbf{y_{mis}}, \theta)]$$

It's now possible to maximize $Q(\theta)$ by choosing the best θ via ML. By iterating back and forth between computing the expectation of $l_{mis}(\cdot)$ and maximizing $Q(\theta)$, our estimate will eventually iterate to a fixed point. This is the well-known EM Algorithm. Dempster, Laird, and Rubin (1977) prove several properties about EM, namely that it consistently converges to the ML value, and that it does so monotonically (so that each iteration between computing the expectation of the missing data contribution and the complete data likelihood increases the likelihood function). We typically superscript EM iterations with k. To compute the next iteration of the EM algorithm $(k+1)$, we use the estimate of θ from the kth iteration, θ^k, to evaluate the expectation of missing data's contribution to the log-likelihood.

The EM algorithm has a long history in the context of multinomial likelihoods (Hartley 1958), because computing the expected contribution of the missing data is straightforward. Once again, the fact that the log-likelihood is linear in the data $\mathbf{y_{jt}}$ and the sufficient statistics $\mathbf{q_{a,x}}$, simplifies the computation dramatically.

$$\begin{aligned} E[l_{mis}(\mathbf{y_{mis}}(\theta^k, \mathbf{y_{obs}})|\theta^{k-1}] &= E\left(\sum_{\forall t \in T_{mis}} \sum_{a \in \{a_t, a_t'\}} \sum_{\forall j \in a} y_{jt,mis} \ln p_j(\theta, a, x_t) | \theta^k\right) \\ &= \sum_{\forall t \in T_{mis}} \sum_{a_t \in \{a_t, a_t'\}} \sum_{\forall j \in a} E(y_{jt,mis}|\theta^k) \ln p_j(\theta, a, x_t) \qquad (9) \end{aligned}$$

The previous equation shows that in order to evaluate the likelihood, we do not need to integrate out over all possible values of the missing data, instead linearity implies we can simply plug in the expected sufficient statistic. We can also define the augmented dataset, $\mathbf{y}_{aug}(\theta^k) = [\hat{\mathbf{y}}_{mis}(\theta^k), \mathbf{y}_{obs}]$, and see that the adjusted likelihood function is the same as the likelihood function which treats the augmented dataset as the complete data (again via linearity). Therefore it is possible to estimate the parameters θ by maximum likelihood.

$$\begin{aligned}
\hat{\theta}^{k+1} &= \arg\max_{\theta} Q(\theta) \\
&= \arg\max_{\theta} l_{obs}(\mathbf{y_{obs}}, \theta) + E[l_{mis}(\mathbf{y_{mis}}, \theta)] \\
&= \arg\max_{\theta} \sum_{t=1}^{T} l(\theta|\mathbf{y}_{aug}(\theta^k)) \qquad (10)
\end{aligned}$$

(This is the M-Step.)

At this point, we must derive an expression for the expectation of the missing data. We consider the case where in period t product k stocks out, and assume that M_t potential consumers are in our market in period t so that $m_t \leq M_t$ of them face choice set a_t and $M_t - m_t$ of them face choice set a'_t. For notational convenience we let $\alpha_t = \frac{m_t}{M_t}$ be the fraction of consumers not affected by the stockout. It becomes clear that for some product $j \neq l$ that conditional on y_{jt} the overall sales observed for that time period, the sales are distributed binomially across the two regimes.[9] In fact the number of sales before and after the stockout (for a single period) are:

$$\begin{aligned}
q_{j,a,x_t,t} &\sim Bin\left(y_{jt}, \frac{\alpha_t p_j(\theta, a_t, x_t)}{\alpha_t p_j(\theta, a, x_t) + (1-\alpha_t) p_j(\theta, a', x_t)}\right) \\
E[q_{j,a,x_t,t}] &= y_{jt} \frac{\alpha_t p_j(\theta, a_t, x_t)}{\alpha_t p_j(\theta, a, x_t) + (1-\alpha_t) p_j(\theta, a', x_t)} \qquad (11) \\
E[q_{j,a',x_t,t}] &= y_{jt} - E[q_{j,a,x_t}] \qquad (12)
\end{aligned}$$

The last two expressions represent the contribution of the period t to the sufficient statistic $\mathbf{q_{a.x}}$. Given values of θ, and α_t it is easy to compute these expectations. We've already discussed using how to update θ^k iteratively, so all that remains is to address α_t. We don't know the true value of α_t, and there are typically three ways to handle this. The standard E-M approach is to specify a distribution for α_t and integrate it out. Another approach is to add an additional step to our iterative procedure where we draw α_t from its marginal distribution, and treat it as data for the E- and M-steps (this is akin to a Gibbs Sampler/MCMC). The third approach is to treat α_t not as missing data, but rather as a parameter to be estimated, and choose the value of α_t which best improves the likelihood. All three of these methods are going to have tradeoffs between how many E-M iterations are required for convergence and how long each iteration takes.[10]

[9] It is important to highlight that this is not an assumption. This follows representationally, from Assumptions 1-3 and our underlying multinomial DGP.

[10] For a fully Bayesian MCMC approach (such as the Gibbs Sampler) we won't actually reach a fixed point

The standard E-M approach is to integrate out the α_t's. Notice this is much easier than integrating the entire likelihood, we only need to integrate to compute the (single dimensional) expectation in (11). In order to do this we must specify a distribution for α_t. Consider m_t is the number of consumers facing choice set a_t which must be determined by the product l that stocks out. We can look at this as the number of consumers before q_{l,a,x_t} bernoulli trials are successful given it must be less than M_t consumers. This is the definition of the conditional negative binomial.[11] Therefore we can write,

$$\frac{\alpha_t}{M_t} = m_t \sim \frac{NegBin(q_{l,a,x_t}, p_l(\theta, a, x_t))}{NegBinCDF(M_t, p_l(\theta, a, x_t))} = w\left(m_t | q_{l,a,x_t}, M_t, p_l(\theta, a, x_t)\right)$$

Since this is a discrete distribution we can compute the expectation over all $m_t \leq M_t$ using the $w(\alpha_t|(\cdot))$ as p.m.f. weights by taking the sum

$$E[q_{j,a,x_t}] = \sum_{m_t=0}^{M_t} y_{jt} \frac{m_t p_j(\theta, a_t, x_t)}{m_t p_j(\theta, a, x_t) + (1 - m_t) p_j(\theta, a', x_t)} w(m_t|(\cdot)) \tag{13}$$

(This is the E-Step.)[12]

The other way to handle α_t is to exploit the duality in the Bayesian world between parameters and missing data to estimate α_t. We can consider the partial likelihood of each period $l_t(\cdot)$, and choose the α_t which maximizes it's contribution to the likelihood function. This is particularly convenient because α_t does not effect the partial likelihood of other periods $l_s(\cdot)$ for $s \neq t$, likewise α_t only enters the $E[y_{jt}]$ and combinatorial terms of the likelihood, and never any terms involving θ. That means we can divide our parameter space and condition out on α_t when searching for θ and vice versa. Even though there are many α_t's to find now (one for each period with missing data), the computational burden is not that bad because it is sufficient to do several hundred single dimensional optimizations which is orders of magnitude easier than a single large dimensional optimization problem. This algorithm typically takes more time per iteration than the E-step estimator above, but fewer iterations to achieve convergence. When such separation of parameter space is possible this is a variant of the EM algorithm often referred to as ECM (Liu and Rubin 1994), and has natural analogues to fully Bayesian (MCMC) approaches.
(This is the C-Step).[13]

By alternating between imputing the missing data (E-Step) and maximizing the (log)likelihood function (M-Step) it is possible to obtain estimates for parameters (θ, α) which increase likelihood monotonically until reaching a fixed point.

but rather a stationary distribution.

[11] This is once again not an assumption on the process but rather a consequence of the multinomial DGP.

[12] The multiple stockout case proceeds analogously, although it relies on the multivariate generalization of the negative binomial, the negative multinomial, and involves more than one sum. The approach can be extended to deal with many unobserved stockouts, the technical details of which are presented in the Appendix.

[13] This was explored in response to a comment Jack Porter made about a previous version.

5 Estimation

5.1 Parametrizations

All that remains is to specify a functional form for $p_j(\alpha, a_t, x_t)$. In this section we'll 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 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. It highlights issues of identification which we will address later.

A typical solution in the differentiated products literature to handling these sorts of problems is to write down a random coefficients logit form for choice probabilities. This still has considerable flexibility for representing substitution patterns, but avoids estimating an unrestricted covariance matrix. What's also nice is that this family of models is 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:

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

Where δ_j is the mean utility for product j, μ_{ij} is the individual specific taste, and ε_{ijt} is the idiosynchratic 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 functional 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). 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 ($\mu_{ij} = 0$) and 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 BLP, $\mu_{ijt} = \sum_l \sigma_l x_{jl} \nu_{il}$, 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 there is sufficient variation in the choice set, Nevo (2000a) shows that product dummies may be used to parameterize the δ_{jt}'s. When we include product dummies it allows us to rewrite the δ's as:

$$
\begin{aligned}
\delta_{jt} &= \underbrace{X_j\beta + \xi_j} + \xi_t + \Delta\xi_{jt} \\
&= \quad\; d_j \quad\; + \xi_t + \Delta\xi_{jt}
\end{aligned}
\tag{14}
$$

where d_j functions as the product "fixed-effect". If we have enough observations, we can also include market specific effects to capture the ξ_t. This changes the interpretation of the structural error ξ, which is traditionally the unobservable quality of the good. The remaining error term $\Delta\xi_{jt}$ is the market specific deviation from the mean utility. Nevo (2000a) points out that the primary advantage is that we no longer need to worry about the price endogeneity and choice of instruments inside our optimization routine, while the mean tastes for characteristics β (along with the ξ_j's) can be captured by a second-stage regression of d_j on X_j.

This highlights some important consequences for our study of stockouts. We expect products which stock out to have higher than average ξ_{jt}'s. Therefore, if we discard data from periods where stockouts occur, this is akin to violating the moment condition on $\Delta\xi_{jt}$ as we are more likely to discard data from the right side of the distribution than the left.[14] Likewise, if we estimate assuming full availability, stocked out products should have large negative $\Delta\xi_{jt}$'s (since they have no sales at all), which once again violates the moment condition on $\Delta\xi_{jt}$.

5.2 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 our dataset are short, it is likely that choice probabilities may vary substantially over periods. Over long periods of time (such as annual aggregate

[14]This correlation may not be as strong as one might expect because stockouts are also correlated with low inventory, which should be uncorrelated (perhaps even slightly negatively) with $\Delta\xi_{jt}$.

data) these variations get averaged out. The distribution of tastes over a long period is essentially the combination of many short-term taste distributions. We only observe what we estimate, so usually that is 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. We might think that x_t includes information such as the time of day, day of the week, or location identifiers. Depending on how finely data are observed, not accounting for this additional heterogeneity may place a priori unreasonable restrictions on data.

We can model this dependence on x_t in several ways. One is to treat $p(\cdot|x_t)$ as a different function for each x_t. In other words we could think about each location having its own distribution of tastes, and parameters θ. We could also imagine a scenario where all markets faced the same distribution of consumers, but that distribution varied depending on the day of the week. In this approach x_t can be thought of as the demographic covariates used in the differentiated products literature (Petrin 2002, Nevo 2000a), or as consumer level micro-data (Berry, Levinsohn, and Pakes 2004), but need not be limited as such. Another way we could think about the x_t are as characteristics of consumers. Some of the parameters in θ might be fixed over x_t's while others depend on x_t. A good example might be to think that the correlation of tastes is constant across all populations but the mean levels are different. When we present the estimates we explore several different such dimensions of heterogeneity.

The other approach we can take is to parameterize M, based on information similar to the information we've incorporated in the x_t's. Thus instead of letting the choice probabilities vary, we could let the number of consumers passing by the machine vary. This becomes helpful because M is going to be the driving force behind substitution to the outside good. It also allows for a common shock across periods without affecting the choice probabilities. In markets with retail data, this can be extremely useful as a way of adjusting for seasonality, holidays, or other events which might have an affect on the effective size of the market. Parameterizing M has a long history in the literature (Berry 1992). In practice, it's pretty easy, and can be considered as a (C-Step). At each iteration we simply find the parameter values for M which maximize the likelihood conditional on the (θ, α). We don't worry about simultaneity because the likelihood factors in M. Once again this approach begins to resemble a fully Bayesian MCMC approach. We present some specifications for M with the estimates.

5.3 Identification of Discrete Choice Models

In this section we address non-parametric and parametric identification of the choice probabilities $p_j(\theta, a_t, x_t)$, while still continuing to assume the underlying d.g.p. 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. For the quite general formal results, the standard reference is Matzkin (1992).[15]

[15] Athey and Imbens (2006) 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.

Nonparametric identification is easily addressed by our $q_{a,x}$ sufficient statistic representation. For a given (a, x), the sufficient statistics must be observable, moreover the efficiency is roughly going to go as $\sqrt{n_{a,x}}$ where $n_{a,x}$ are the total number of consumers facing (a, x). 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 most likely not be nonparametrically identified.

Typically we use the random coefficients parameterization presented above, so we're more worried about whether that is going to be parametrically identified. One approach might be to assume a smooth functional form for $p_j(\cdot)$ and then use delta method arguments to do a change of variables to the parameterized version, but a heuristic sufficient statistics based argument may be easier to understand (and hopefully more useful) for the applied researcher.

The typical source of identification in the differentiated products literature is by long run variation in the choice set. For this to be useful as a source of identification, these variations must be exogenous to the model. Thus we could think about each "observation" as being the $q_{a,x}$ sufficient statistics we've presented in this paper. The way to think about these models is to compute the effective number of $q_{a,x}$ "observations" and compare them to the parameters we're hoping to explain. We can see this by constructing a matrix of observables which describes each $q_{a,x}$. In many discrete choice models these may be characteristics, product dummies, time dummies, of the available products. For nonlinear effects, (tastes for example), interaction terms should also be included. If we want to see if we can estimate all of those parameters, we could think about determining whether or not our "data" matrix is of full-column rank. We also see that many of our observations will have the same descriptive variables, and thus will be linearly dependent. Once again the $q_{a,x}$ representation makes this quite clear, as we don't have additional observations but rather our additional observations get added into the sum of $q_{a,x}$ (which may improve efficiency but not allow us to identify additional parameters). Stockouts are useful, particularly when trying to identify product dummies, because they provide linearly independent observations of $q_{a,x}$. If we only ever observe a change from choice set $a \to a'$ (suppose the only product that ever stocks out is Snickers), then we only have two effective "observations". If we observe lots of stockouts and different choice sets, then we have the potential to observe $J \times (J - 1)$ "observations".

In Berry, Levinsohn, and Pakes (1995) and related literature the choice set is not a collection of products but rather a collection of bundles of characteristics. Thus their 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 characteristics $a_t = \{\forall j \in t : \mathbf{z_j}\}$ (including price), for each product. This is important because the primary source of identification comes from variations in one of the characteristics, price, across time and markets. When price changes from one period to the next, this represents a change in the a_t, albeit usually only along a single dimension of $\mathbf{z_j}$. Other sources of variation in the choice set involve changes in product characteristics as they vary from year to year (mileage, HP, etc.). The third source of choice set variation is when new goods are introduced, and an entirely new $\mathbf{z_j}$ is provided to consumers. A possible disadvantage to applying the technique of BLP to other industries is that there may not be sufficient variation in the characteristics of products from one year to the next, or that variation (changes in characteristics and product mix) is often endogenously determined by the participants.

Our model presents a different way to interpret variation in choice sets. In our context a_t doesn't vary with long term product mix, or potentially endogenous pricing decisions, but rather as products stock out. When products stock out they are no longer in the set of available products a_t. This variation is potentially more useful because our new choice set does not necessarily look like the old choice set with a single dimension of $\mathbf{z_j}$ altered. Instead, we add and remove entire $\mathbf{z_j}$'s similar to the new products case. This is particularly helpful, because it allows us to observe substitution not on a single characteristic at a time, but jointly over several characteristics. Moreover, when stockouts happen one at a time, we know which joint distribution to attribute those changes to. Using long-term variation, product mixes vary roughly simultaneously from model year to model year.

Additionally, this variation is 100% exogenous as we've written down our model, 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. While firms can restock the machine (or even change the product mix to prevent future stockouts), this does not change the fact that any particular stockout is exogenous to the model.

Finally, one of the most common applications of these sorts of models is to predict substitution probabilities. It should be clear that 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.

5.4 Computation

So far, the M-step has been presented in the context of maximum likelihood (ML), or in the case where random coefficients are used maximum simulated likelihood (MSL). The algorithm for these approaches would be to simulate consumers, compute the choice probabilities, and evaluate the likelihood.

It is important to recognize that by using the EM algorithm or one of its variants we are not locked into ML estimation. A typical approach to the demand side in a differentiated products setting is to iterate on the contraction mapping from Berry, Levinsohn, and Pakes (1995). This is particularly effective because it solves for the unique set of optimal δ's monotonically without computing anything other than choice probabilities. More importantly the fit is exact and unique. A further advantage of GMM approaches over ML is the ability to incorporate additional (structural) moments along with the demand side.[16]

A GMM approach using demand-side moments can be used in lieu of ML estimation in the M-step. In fact, the Generalized EM algorithm (GEM) Dempster, Laird, and Rubin (1977) says that any approach which improves the log-likelihood at each EM iteration (rather than fully maximizing the likelihood of the augmented dataset) will converge to the correct

[16] Draganska and Jain (2004) develop a method for incorporating supply side information in the ML framework which appears to be agreeable with the missing data procedure we've provided.

value. In other words it is sufficient that θ is any sequence which obeys:

$$l\left(\theta^{(k+1)}, \mathbf{y}_{aug}(\theta^{(k)})\right) \geq l\left(\theta^{k}, \mathbf{y}_{aug}(\theta^{(k-1)})\right) \tag{15}$$

should lead to consistent EM corrected estimates.

Therefore, we can replace the ML estimator in the M-step, with the easier to compute GMM estimator from Berry, Levinsohn, and Pakes (1995) as long as we verify that the condition (15) holds at each step, it is perfectly acceptable to use δ's obtained from the contraction mapping of BLP rather than obtaining ML estimates. Typically the contraction mapping in Berry, Levinsohn, and Pakes (1995) is written as

$$\delta_j^{(n+1)} = \delta_j^{(n)} + \ln(s_j) - \ln p_j(\theta, a_t, x_t)$$

However, $E[\ln p_j(\theta, a_t, x_t)]$ is not linear in the missing data, and computing could be tricky. This is easily remedied as the standard computational trick is to use the exponentiated δ's in the contraction mapping and then we have that:

$$\exp(\delta_j^{(n+1)}) = \exp(\delta_j^{(n)}) \frac{q_j}{\hat{q}_{j,a_t,x_t}(\theta)}$$

While optimizing the likelihood function can be expensive (particularly if simulation is required), evaluating is relatively cheap. Once this condition fails to hold, a switch to ML estimates should be made just to check convergence, but in most cases an additional ML iteration does not improve the overall likelihood.[17]

A useful way to think about what the EM adjustment does is to note that there are some missing data regarding which products are available. Without knowledge that matches sales to the sets of available products, the values for δ are incorrect. The EM algorithm uses observable information about the substitution patterns to replace unknown δ's with more accurate estimates, and then recomputes estimates for the parameters. All other algorithms (GMM, ML, MH, Gibbs Sampling) are simply ways to get to the maximum. For the results we report later, we use GMM. Details for the exact moment conditions used are contained in the Appendix.

[17] This is a result of the exact fit of the δ's in the BLP moments, and the fact that we don't include supply side moments. If supply side moments are included we don't expect ML and GMM to give the same estimates. In this scenario, the EM-GMM estimates are probably appropriate. The problem is that we aren't aware of a broad theoretical result for incorporating missing data in moment condition models as Dempster, Laird, and Rubin (1977) do for ML.

6 Industry Description, Data, and Reduced-form Results

6.1 The Vending Industry

The vending industry is well suited to studying the effects of product availability in many respects. For one, product availability is well defined. Products 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 uniform across all of the machines in the 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.[18]

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 also 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 machine. Sometimes the client specifies a minimum number of machines and locations, and several optional machines and locations.

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

In order to measure the effects of stock-outs, we use data from 58 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 February-June 2003, which corresponds with the spring term at ASU.

Each of these machines collects Digital Exchange (DEX) data. DEX is the vending industry standard data format, and 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. The data are typically transferred to a hand-held device by the route driver while he services and restocks the machine. This device is then synchronized with a computer at the end of each day. In our dataset, (thought not typically) additional inventory observations are made between service visits, because DEX data are wirelessly transmitted several times each day. As of 2003, the ASU route was the only route to be fully wireless

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

enabled.

6.2 Data Description

The dataset consists of snack and coffee machines; we focus on the snack machines in this study. Throughout the period of observation, the machines stock around 70 different products, including chips, crackers, candy bars, packaged donuts, gum, and mints. 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 [19] or have insubstantial sales (usually less than a dozen total over all machines). In the examples we present, we exclude these items in addition to excluding gum and mints, based on the assumption that these products are substantially different from more typical snack foods, and rarely experience stockouts. Including gum and mints does not substantially change our results. It is important to note that not every product appears in every machine. The 50 products in the dataset are listed in Tables 1 and 2.

Typically, sales are only observed when vending machines are refilled. Thus in order to have data before and after a change in product availability occurs, "perpetual" data collection would be required. The data from Arizona State University are interesting because periodic wireless readings of the inventory data are observed each day (often several times). This provides two distinct advantages: the observation of the machine is no longer linked to the restocking of the machine,[20] and the machine's inventory is sampled more frequently. These help to mitigate the limitations of the periodic inventory system. The methodological goal of this paper is to provide consistent estimates of demand not only for perpetual inventory systems but for periodic ones as well.

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. This is somewhat similar to the approach taken by Nevo (2000a) for RTE cereal.

The retail prices observed in the vending machine are constant over time and across broad groups of products as shown in Tables 1 and 2. Baked goods typically vend for \$1.00, chips for \$0.90, cookies for \$0.75, candy bars for \$0.65, and gum and mints for \$0.60. This makes for a simpler and less complicated framework for static models of demand. As compared to typical studies of retail demand and inventories (which often utilize supermarket scanner data), there are no promotions or dynamic price changes (Aguirregabiria 1999). This presents a bit of a problem, because for the most part prices do not vary within a particular product category. 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

[19] While often sold alongside of snacks in vending machines, condoms are poor substitutes for potato chips, and are not included in our sample.

[20] This is not exactly true. While a wireless observation can be made without restocking the machine, the wireless readings are also available to the vending company, and thus decisions to refill are endogenous. For a static analysis of stock-outs that is not concerned with the retailer's dynamic restocking problem, this is not problematic.

example this is not the case.

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. 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 products are listed in Tables 1 and 2. By examining Tables 1 and 2, several trends become apparent. There is a lot of variation in the markups of the products. Markups are lowest on branded candy bars (about 50%), and markups are highest on the Big Grab^{TM}chips (about 70%). The product with the highest markup is the Peter Pan crackers, which have an average markup of nearly 82%. Table 3 reports regression estimates from a regression of markup on the percentage of the time a product is stocked out. The result shows what one might expect; namely that products with high markups are less likely to stock out than products with lower markups. This is also true when product category is adjusted for.

Other costs of holding inventory are also observed, including the spoilage and number of products removed from machines. 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 (approximately 2 weeks) than most products, which may last several months before spoiling. For this static analysis of demand, we assume that the costs associated with spoilage are negligible.

6.3 Reduced-form Results

Before applying the estimation procedure described above to the dataset, first consider a simple reduced form analysis of stockouts. Table 4 reports the results of a regression of stockout rates on starting inventory levels. We report results for Probit and OLS with and without product fixed effects. What we find is not surprising, namely 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 potato chips usually contains 14 units. This means that the OLS (fixed effects) probability of witnessing a stockout from a full machine in 3-day period is $.242 - .008 * 14 = 13\%$. For a machine with a starting inventory of only two units, the predicted chances of a stockout are one in four.

In table 5, we compute the average hourly profits for each four hour wireless time period. Then, we regress that on the number of products stocked out. The first specification (Column 1) estimates the hourly cost to be about $0.05 per product stocked out. Since the number of products stocked out across the entire machine might not matter as much as the number of products stocked out in each category, we include category by category stockouts in Column 4. These estimate the costs per stockout at around $0.11 per Big Grab bag of chips to $0.55 per bag snack. Column 2 examines the effect of a stockout in the category with the most stockouts and estimates this effect to be about $0.66 per hour. Columns 3 and 5 also include indicators for the number of products stocked out in the category with the most stockouts. 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 even faster.

7 Empirical Results

For the discrete choice model, several different specifications are addressed. The logit, nested logit, and random coefficients logit models are estimated with the assumption that the missing data are ignored. The nested logit and random coefficients model are also estimated with the proposed correction for missing data. Finally, the random coefficients model is also estimated under the assumption of full product availability. An aside, that should be pointed out is that there is no missing data corrected logit model. The IIA property of the logit model implies that the missing data is perfectly ignorable. In fact, removing a product from the product mix and re-estimating is a typical specification test for the standard logit model.

There are a number of ways in which we could condition on observable characteristics. We could run everything machine-by-machine, or pool the data from different machines. [Add discussion of the trade-offs of different conditioning decisions in estimation, and results of robustness tests...] The results that follow pool across machines case, the linear term is a product dummy (d_j) and M_t is modelled as a machine fixed effect. We include all observable characteristics in the nonlinear part.

Table 7 reports the corrected values of many of the product dummies (the d_j's) under each of the different models.[21] As most products experience stock-outs, the correction for stock-out events tends to increase the estimate of the d_j's for each product. Naturally, if a product experiences stock-outs only when other products are stocked-out, this correction can go the other way (i.e., the bias from forced substitution exceeds the bias from censoring), as we see in the case of Peter Pan Crackers. (More on this to come....)

The estimated values of the non-linear paramaters (σ's), and the results of the second-stage regression of the d_j's on characteristics (including the R^2 from this regression) are reported in Table 6. Getting the δ_{jt}'s and hence d_j's corrected means substantially different estimates from the ignore missing data and full availability case for the mean taste parameters.

Tables 8, 9, 10, 11 and 12 are all currently from the same representative machine. We are re-running some of these with more to come soon.

Tables 13 and 14 present robustness tests and comparisons of different specifications for the heterogeneity across machines and time periods. Table 13 examines heterogeneity where estimates are conditional x_t, and Table 14 presents the effects of parametrizing M. In each table the mean square error (MSE) is computed and the percentage of variation explained as compared to an unrestricted model is also reported. The interpretation is meant to be

[21] A few d_j's were excluded just to keep the table intact, but they follow the same pattern.

similar to an R^2 which obviously doesn't exist for this model.

8 Counterfactual Experiments

These estimates are now used to predict the effect that stockouts have on the profits of the vending operator. For simplicity, the model was estimated using data from a single snack machine located in Alumni Hall on the campus of Arizona State University. This machine was chosen because it was relatively high sales volume, and was not located particularly close to the other machines in the dataset. Two different availability regimes were compared, under the first availability regime (labeled A), it was assumed that the product most likely to stockout from each category was stocked out. Under the second availability regime (labeled B), three popular types of chips: Big Grab Doritos, Big Grab Rold Gold Pretzels, and Big Grab Sunchips Harvest Cheddar were assumed to have stocked out. By comparing these availability regimes with a full availability regime, it is possible to computed expected loss to the producer in $(p-c)$ terms. Furthermore, it is estimated that approximately 382 consumers "consider" a purchase from the vending machine each hour[22] though many choose to purchase no product at all. Simulations are performed to obtain estimated sales under regime A and B, and they are compared to simulations of sales at the full machine using the specifications for demand in A,B, and the full machine. Table 11 reports the expected lost profits when compared to the full machine with the same specification for demand. The results in Table 11 indicate that the specification which assumes full availability always predicts the lowest profit loss, and the EM corrected model predicts the most, and twice as much as the full availability specification. When these quantities are compared with accounting data on the amount of cash collected from the machine (about \$150/week) the loss from stockouts is substantial. Table 12 examines stockout regime B under different numbers of consumers, in an attempt to demonstrate the different curvature of the cost of stockouts. (It is important to remember that many consumers still prefer the outside good).

Tables 13 and 14 examine the variation in the marketsize M_t and the observed sales q_{jt}, to provide some insight as to how to model heterogeneity in our empirical example. We used a nonparametric approach (conditional means) to predict each cell in the dataset. We reported the mean-squared-error of the predictions, as well as the adjusted (for parameters and degrees of freedom) R^2 and the implicit number of parameters involved in using that approach. We found that simply accounting for the time of day, or day of week with 6 or 7 dummy variables respectively did an excellent job explaining the overall variation in the size of the market. This indicates that the overall size of the market varies more across time than across machine locations.

Likewise, the inclusion of product dummies accounted for nearly 99% of the variation in the observed sales. Controlling for machine location, or time specific effects added complexity without improving the result. This indicates that sales of individual products depend mostly on the product's identity, and much less on which machine it is in or what time of day it is. This indicates that we can probably pool our high frequency data across time (and locations) without worrying about ignoring some important source of heterogeneity.

[22] between 12pm-4pm on weekdays

9 Conclusion

This paper has demonstrated that failing to account for product availability correctly can lead to biased estimates of demand, and that these biased estimates can lead to economically meaningful results when trying to measure the welfare costs of stockouts. Rather than examining the effect of changing market structure (entry, exit, new goods, mergers, etc.) on market equilibrium outcomes, we seek to understand the effect that temporary changes to the consumer's choice set have on producer profits (and our estimators). The differentiated products literature in Industrial Organization has used long term variations in the choice set as an important source of identification for substitution patterns, 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. Finally, the welfare impact of stockouts in vending machines has been shown to have a substantial effect on firm profits indicating that product availability may be an important strategic and operational concern facing firms and driving investment decisions.

A Estimation Details and the Case of Multiple Stockouts

A.1 Estimation Details

Following the specification described in Nevo (2000b) the observables are broken into two categories. The first category, X_1, contains the the price, and the product dummies. The second category, X_2, contains a constant, the price, and the product characteristics (calories, sugar, etc.). This way the product dummies are able to absorb ξ_j (the unobserved characteristics), while the substitution patterns are driven by random tastes for the product characteristics. Now consider the mean utility (linear part) as:

$$\begin{aligned} \delta_{jt} &= x_{1j}\beta - \alpha p_{jt} + \xi_j + \xi_t + \Delta\xi_{jt} \\ \delta_{jt} &= d_j + \xi_t - \alpha p_{jt} + \Delta\xi_{jt} \end{aligned} \tag{16}$$

The nonlinear part of the utility (random part) is represented by μ_{ijt} as defined below:

$$\mu_{ijt} = \sum_k x_{2jkt}\sigma_k\nu_{ikt} + \epsilon_{ijt}. \tag{17}$$

By combining (16) and (17) we obtain the overall utility for each consumer as:

$$u_{ijt} = \delta_{jt} + \mu_{ijt}. \tag{18}$$

The marketshares for the random coefficients logit take on the form (Berry, Levinsohn, and Pakes 1995)

$$s_{jt} = \int \frac{\delta_{jt} + \sum_k x_{2jkt}\sigma_k\nu_{ikt}}{1 + \sum_j \delta_{jt} + \sum_k x_{2jkt}\sigma_k\nu_{ikt}} f(\nu)d\nu \tag{19}$$

Unfortunately this integral does not have a closed form, therefore it must be approximated using a simulation estimator (Berry, Levinsohn, and Pakes 1995) by taking ns draws of the ν_{ik} variables where $ns = 1000$.

$$\hat{s_{jt}}(\delta;\cdot) = \frac{1}{ns}\sum_{i=1}^{ns} s_{ijt} = \frac{1}{ns}\sum_{i=1}^{ns} \frac{\delta_{jt} + \sum_k x_{2jkt}\sigma_k\nu_{ikt}}{1 + \sum_j(\delta_{jt} + \sum_k x_{2jkt}\sigma_k\nu_{ikt})} \tag{20}$$

Now define the vector of predicted marketshares as $\hat{S}_t(\delta;\cdot) = [\hat{s_{1t}}, \hat{s_{2t}}, \ldots \hat{s_{Jt}}]$ and the vector of observed marketshares as S.

$$\delta_t^{h+1} = \delta_t^h + \ln[S_t] - \ln[\hat{S}_t(\delta^h;\cdot)] \tag{21}$$

Berry, Levinsohn, and Pakes (1995) show that the mean values δ_{jt} can be computed by iterating the contraction mapping on the vectors $S_t, \hat{S}_t(\delta^h;\cdot), \delta_t^h$ until $\|\delta_t^{h+1} - \delta_t^h\| < \varepsilon$ or the difference between the mean utility levels is arbitrarily small.

After estimating the mean utility levels, the error term (which defines the moment conditions) is simply:

$$\xi = \delta(S_t; \theta) - d_j + \alpha p_{jt} = \omega(\theta). \tag{22}$$

Here the subscript on the ξ error term is left intentionally ambiguous, and the possible ways of defining this error term are discussed in the estimation section.

The parameters can no longer be obtained by least squares regression, so a nonlinear search must be performed instead. We use a two-step instrumental variables GMM procedure (Hansen 1982) as is typical in these models, with Z as instruments and weighting matrix $A = I$ in the first step and $A = E[Z'\omega\omega'Z]$ in the second step. The minimization problem becomes:

$$\hat{\theta} = \min_{\theta} \omega(\theta)' Z A^{-1} Z' \omega(\theta) \tag{23}$$

A.2 Multiple Unobserved Stockouts

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

Let's suppose that we have two products which stockout in period t. We'll label those products B and A, if we do not observe the timing of the stockouts, then there are four possible inventory regimes. The inventory regime with full availability, the regime with only A stocked out, the regime with only B stocked out, and the regime where both A and B are stocked out. We denote these availability sets (a_0, a_A, a_B, a_{AB}). Now if for product j we observe y_{jt} sales in period t then the expected number of sales to have occurred in each regime is:

$$E[q_{jti}] = y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)}$$

The only unknown element to compute that expectation is the α's. The approach is the same as before (to integrate them out). The only problem is that for a single stockout α was two dimensional and could be represented by a single parameter (since the other was just $1-\alpha$). Now α is four dimensional (three parameters). For the n stockout case, there are 2^n values of α to impute, which implies $2^n - 1$ parameters.

$$
\begin{aligned}
E[q_{jti}] &= \sum_{\forall \alpha_A, \alpha_B, \alpha_{AB}, \alpha_0 : \alpha_0 + \alpha_A + \alpha_B + \alpha_{AB} = 1} y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)} g(\alpha_A, \alpha_B, \alpha_{AB}, \alpha_0 | \theta, y_{jt}) \\
E[q_{jti}] &= \sum_{\forall \alpha_i : \sum \alpha_i = 1} y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)} g(\hat{\alpha} | \theta, y_{jt})
\end{aligned}
$$

where $\hat{\alpha}_l = [\alpha_0, \alpha_A, \alpha_B, \ldots, \alpha_{AB}, \ldots]$ is a vector of the appropriate 2^n α values.

All that remains is to write down the joint distribution $g(\hat{\alpha}|\theta, y_{jt})$. We show how to construct the joint density $g(\cdot)$ for the two stockout case, but it should be clear that this approach can be easily extended to construct the joint density for the n stockout case. There are two possible sequences of availability regimes $R : a_0 \to a_A \to a_{AB}$ or $S : a_0 \to a_B \to a_{AB}$. We can affix a probability to each sequence (with some abuse of notation), we define $z_R = Pr(a_0 \to a_A \to a_{AB})$ and $z_S = Pr(a_0 \to a_B \to a_{AB})$. It happens here that $z_A = 1 - z_B$ but everything we've written can be extended to n stockouts.

Now let's condition on the assumption that event R actually took place, we write $m_{A,R,t} = \alpha_{A,R} M_t$ for convenience, we'll drop the t subscripts and focus only on a single time period. Then we write $m_{0,R} = \alpha_{0,R} M$, $m_{A,R} = \alpha_{A,R} M$ $m_{AB,R} = \alpha_{AB,R} M$ and as the number of consumers that would have faced regimes a_0, a_A, a_{AB} respectively if event R had occurred. We also need to define the beginning of period inventories $\omega_{\mathbf{t}} = [\omega_{At}, \omega_{Bt}, \ldots]$. Once again for convenience we drop t subscripts. With everything now defined, we can write down the probabilities conditional on R.

$$
\begin{aligned}
& Pr(M_{0,R} = m_{0,R}, M_{A,R} = m_{A,R}, M_{AB,R} = m_{AB,R} | X_A = \omega_A, X_B = \omega_B) \\
= \; & Pr(m_{0,R} | X_A = \omega_A, X_B = x_{B,0}) \cdot Pr(m_{A,R} | X_{B,A} = \omega_B - x_{B,0}) \cdot Pr(m_{AB,R} = M - m_{A,R} - m_{0,R})
\end{aligned}
$$

There are three parts. The third part is trivial, the probability is one so long as $M \geq m_{A,R} + m_{0,R}$ and zero otherwise. The second is the negative binomial, and the first is the negative multinomial. We can rewrite as follows:

$$
\begin{aligned}
Pr(m_{0,R}, m_{A,R}, m_{AB,R}, R) &= Pr(m_{0,R} | X_A = \omega_A, X_B = x_{B,0}) \cdot Pr(m_{A,R} | X_{B,A} = \omega_B - x_{B,0}) \\
Pr(m_{0,S}, m_{B,S}, m_{AB,S}, S) &= Pr(m_{0,S} | X_B = \omega_B, X_A = x_{A,0}) \cdot Pr(m_{B,S} | X_{A,B} = \omega_A - x_{A,0})
\end{aligned}
$$

Because R and S have be constructed as mutually exclusive events we can add their probabilities.

$$
\begin{aligned}
h(m_0, m_A, m_{AB}, m_B) &= Pr(m_{0,R} | X_A = \omega_A, X_B = x_{B,0}) \cdot Pr(m_{A,R} | X_{B,A} = \omega_B - x_{B,0}) \\
+ & \quad Pr(m_{0,S} | X_B = \omega_B, X_A = x_{A,0}) \cdot Pr(m_{B,S} | X_{A,B} = \omega_A - x_{A,0})
\end{aligned}
$$

We also require that $h(\cdot) = 0$ if both $m_A, m_B > 0$. Now we consider the other case S, and put the two together. We've now constructed an unnormalized density for the joint distribution $h(\cdot)$. To normalize we simply sum over all possible values (since the distribution is discrete). Note that the density must equal zero if $m_A, m_B > 0$.

$$\begin{aligned}
H(m_0, m_A, m_{AB}, m_B) &= \\
& \sum_{m_{0,R}=0}^{M} \sum_{m_{A,R}=0}^{M-m_{0,R}} Pr(m_{0,R}|X_A = \omega_A, X_B = x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A} = \omega_B - x_{B,0}) \\
&+ \sum_{m_{0,S}=0}^{M} \sum_{m_{B,S}=0}^{M-m_{0,S}} Pr(m_{0,S}|X_B = \omega_B, X_A = x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B} = \omega_A - x_{A,0}) \\
= \sum_{m_0=0}^{M} \sum_{m_A=0}^{M-m_0} \sum_{m_B=0}^{M-m_0} \sum_{m_{AB}=0}^{M-m_A-m_B} & Pr(m_{0,R}|X_A = \omega_A, X_B = x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A} = \omega_B - x_{B,0}) \\
&+ Pr(m_{0,S}|X_B = \omega_B, X_A = x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B} = \omega_A - x_{A,0})
\end{aligned}$$

Now we can define $g(\cdot)$ as:

$$g(\alpha_0, \alpha_A, \alpha_{AB}, \alpha_B) = \frac{h(m_0, m_A, m_{AB}, m_B)}{H(m_0, m_A, m_{AB}, m_B)}$$

Finally we can construct the expectation:

$$E[q_{jti}] = \sum_{m_0=0}^{M} \sum_{m_A=0}^{M-m_0} \sum_{m_B=0}^{M-m_0} \sum_{m_{AB}=0}^{M-m_A-m_B} y_{jt} \frac{\alpha_i p_j(\theta, a_i, x_t)}{\sum_{\forall l} \alpha_l p_j(\theta, a_l, x_t)} g(\alpha_0, \alpha_A, \alpha_{AB}, \alpha_B)$$

A.2.1 Negative Multiomial

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.

By definition the negative multinomial tells us:

$$Pr(N = n | X_1 = x_1, X_2 = x_2) = \binom{n-1}{X_1, X_2} p_1^{x_1} p_2^{x_2}$$

When $X_1, X_2, \ldots, X_n \sim Mult(p_1, \ldots, p_n)$.

We are currently researching the PMF and/or characteristic function for this distribution.

A.2.2 Computation

The expectation will have on the order of $O(2^n \cdot M_t)$ elements, where n is the number of stockouts, and M_t is the overall number of consumers in that period. In the ASU vending data we have that $n \leq 4$ always, and less than 1% of observations had $n = 2$ (and only a few observations with more). For $n \leq 10$ and $M_t \leq 1000$ which might constitute a reasonable upper bound for retail environments with daily data observations, there are about one million points, which still takes less than a second to compute on a (pretty old) Pentium 4.

Product	Category	Sales	%Time SO (Lower)	%Time SO (Upper)	p	c	Share	Markup
Gma Oatmeal Raisin	Snack	14805	2.38	3.05	0.75	0.23	2.64	69.33
Chips Ahoy	Snack	13466	0.95	1.15	0.75	0.25	2.4	66.67
Knotts Raspberry	Snack	10053	0.31	0.37	0.75	0.19	1.79	74.67
Gma Choc Chip	Snack	8788	2.93	3.78	0.75	0.22	1.56	70.67
Nutter Butter Bites	Snack	8137	0.17	0.30	0.75	0.25	1.45	66.67
Gma Mini cookie	Snack	5210	19.76	20.38	0.75	0.21	0.93	72
Nutter Butter	Snack	1494	0.51	0.51	0.75	0.27	0.27	64
Rold Gold	Grab	22271	5.60	7.46	0.9	0.27	3.96	70
Sunchip Harvest	Grab	21563	5.89	7.59	0.9	0.25	3.84	72.22
Dorito Nacho	Grab	19624	4.82	7.06	0.90	0.27	3.49	70
Cheeto Crunchy	Grab	19188	5.25	6.94	0.90	0.34	3.42	62.22
Gardettos Snkns	Grab	18005	4.86	6.51	0.75	0.23	3.21	69.33
Ruffles Cheddar	Grab	15062	5.11	6.19	0.90	0.27	2.68	70
Frito Corn Chip	Grab	10887	3.17	4.28	0.90	0.25	1.94	72.22
Lays Potato Chip	Grab	9573	3.33	4.00	0.90	0.17	1.7	81.11
Munchies	Grab	8730	5.28	6.80	0.90	0.25	1.55	72.22
Hot Munchies	Grab	5993	2.58	3.83	0.75	0.25	1.07	66.67
Dorito Guacamole	Grab	4588	3.18	3.91	0.90	0.28	0.82	68.89
Frito Chilli Cheese	Grab	3772	4.14	5.03	0.90	0.28	0.67	68.89
Frito Jalepeno	Grab	2011	12.35	14.47	0.90	0.28	0.36	68.89
Lays Baked Potato	Grab	749	0.89	1.34	0.90	0.49	0.13	45.56
Gardettos	Grab	710	1.60	3.53	0.75	0.3	0.13	60
KC Masterpiece	Grab	406	4.34	4.75	0.90	0.28	0.07	68.89
Ruffles Baked Cheddar	Grab	400	12.00	16.70	0.90	0.28	0.07	68.89
Kettle Jalepeno	Grab	395	7.64	9.14	0.90	0.28	0.07	68.89
Cool Ranch	Grab	129	13.23	15.93	0.90	0.28	0.02	68.89

Table 1: Summary of Products and Markups

Product	Category	Sales	%Time SO (Lower)	%Time SO (Upper)	p	c	Share	Markup
Snickers	Candy (1)	46721	1.32	1.68	0.75	0.33	8.32	56
Twix Caramel	Candy (1)	34250	1.08	1.46	0.75	0.33	6.1	56
M&M Peanut	Candy (1)	26503	1.97	2.54	0.75	0.33	4.72	56
Reese's Cup	Candy (1)	13213	1.14	1.54	0.75	0.33	2.35	56
Kit Kat	Candy (1)	12101	0.93	1.11	0.75	0.33	2.15	56
M&M	Candy (1)	11856	1.30	1.73	0.75	0.33	2.11	56
Caramel Crunch	Candy (1)	11581	0.47	0.83	0.75	0.33	2.06	56
Hershey Almond	Candy (1)	10122	0.89	1.13	0.75	0.33	1.8	56
Crunch Nestle	Candy (1)	340	0.00	0.00	0.75	0.33	0.06	56
Starbursts	Candy (2)	17734	1.08	1.44	0.75	0.33	3.16	56
Kar Nut Sweet & Salty	Candy (2)	16800	2.13	2.79	0.75	0.22	2.99	70.67
Skittles	Candy (2)	10130	1.37	2.04	0.75	0.34	1.8	54.67
Snackwell	Candy (2)	9770	0.79	0.87	0.75	0.28	1.74	62.67
Oreo	Candy (2)	6304	0.23	0.24	0.75	0.22	1.12	70.67
Payday	Candy (2)	5373	0.00	0.00	0.75	0.35	0.96	53.33
Peter Pan (Crck)	Candy (2)	4734	6.34	10.08	0.75	0.12	0.84	84
Peanuts	Candy (2)	4707	1.10	1.21	0.75	0.26	0.84	65.33
Poptart	Pastry	20703	4.20	5.30	1.00	0.35	3.69	65
Banana Nut	Pastry	15793	9.57	12.25	1.00	0.4	2.81	60
Choc Donuts	Pastry	15511	14.89	18.34	1.00	0.46	2.76	54
Ding Dong	Pastry	15468	15.00	19.63	1.00	0.49	2.75	51
Rice Krispies	Pastry	11300	1.94	2.40	1.00	0.31	2.01	69
Pastry	Pastry	3744	11.39	15.60	1.00	0.46	0.67	54
Choc Croissant	Pastry	216	36.19	36.82	1.00	0.38	0.04	62

Table 2: Summary of Products and Markups (cont.)

	Markup	Markup
% Stockout	-0.15	-0.24
(SE)	(0.097)	(0.12)
Constant	66.34	
Snacks		70.87
Grab		71.71
Candy (1)		58.2
Candy (2)		65.36
Pastry		66.98
R^2	.0337	0.7877

Table 3: Regression of Markup on Stockout Rates

	Probit	Probit	OLS	OLS
Constant	-0.368	-1.408	0.251	0.242
Inventory	-0.059	-0.0226	-0.087	-0.008
$\frac{dy}{dx}$	-0.011	-0.004	-	-
Product FE	-	x	-	x
R^2	0.059	0.2323	0.035	0.240

Table 4: Regression of Stockout Rates on Starting Inventory Levels

Profit/Hr	1	2	3	4	5
Constant	1.341	1.391	1.437	1.419	1.283
Stockouts	-0.058	-0.022	-0.069		
(SE)	(0.013)	(0.017)	(0.016)		
Category Max SO		-0.177			
		(0.054)			
Cat Max =2			-0.663		-0.841
			(0.177)		(0.172)
Cat Max =3			-0.227		-0.628
			(0.268)		(0.255)
Cat Max $>$ 3			-0.786		0.057
			(0.304)		(0.255)
Snack SO				-0.548	
				(0.186)	
Grab SO				-0.112	
				(0.051)	
Candy (1) SO				-0.113	
				(0.102)	
Candy (2) SO				-0.329	
				(0.188)	
Pastry SO				-0.467	
				(0.069)	

Table 5: Reduced Form Results for Cost of Stockouts

	Full Avail	Ignore Missing	EM	Nested-Full	Nested-Ignore
constant	-6.7827	-7.4048	-8.0697	-7.4166	-6.7879
calories	18.5995	10.4468	29.3495	10.5818	18.6389
fat	-17.2939	-9.9623	-27.4792	-10.0885	-17.3325
carbs	-6.607	-4.1476	-10.167	-4.1512	-6.6224
sugar	0.6727	1.0333	0.6035	1.0273	0.6866
salt	-1.0614	-0.0828	-1.9868	-0.0676	-1.0417
chocolate	-0.1417	-0.0087	-0.2671	-0.0056	-0.1381
cheese	0.2185	0.2055	0.3032	0.2051	0.222
$\sigma_{calories}$	0.0139	0.0771	0.081		
σ_{fat}	0.1359	0.22	0.24		
σ_{carbs}	0.0283	0.0736	0.0563		
σ_{sugar}	0.0797	0.2041	0.1923		
σ_{salt}	0.1637	0.0399	0.0224		
σ_{choc}	0.0734	0.0666	0.0521		
σ_{cheese}	0.049	0.1407	0.181		
R^2 of Characteristics for d_j	0.2193	0.192	0.2075	0.2194	0.1958
# Obs	155,184	61,844	147,923	155,184	61,844

Table 6: Parameter Estimates

	Ignore	Full Avail	Nested Full Avail	Nested Ignore	EM Rand Coeff
Snickers	1.28	1.09	1.10	1.28	1.28
M&M Peanut	1.11	0.99	0.99	1.11	1.11
Hershey Almond	0.96	0.88	0.88	0.96	0.96
Kit Kat	0.97	0.89	0.89	0.98	0.98
Reese's Cup	1.00	0.91	0.91	1.00	1.00
Twix Caramel	1.18	1.03	1.03	1.18	1.18
Dorito Nacho	1.03	0.92	0.92	1.03	1.03
Sunchip Harvest	1.08	0.95	0.95	1.08	1.08
Cheeto Crunchy	1.09	0.95	0.95	1.09	1.09
Rold Gold	1.07	0.95	0.95	1.07	1.07
Ruffles Cheddar	1.02	0.92	0.92	1.02	1.02
Pastry	1.06	0.93	0.94	1.06	1.06
Poptart	1.08	0.95	0.95	1.08	1.08
Rice Krispies	0.97	0.89	0.89	0.97	0.97
Gma Oatmeal Raisin	1.00	0.91	0.91	1.00	1.00
Nutter Butter Bites	0.93	0.87	0.87	0.93	0.93
Knotts Raspberry	0.94	0.87	0.87	0.94	0.94
Chips Ahoy	0.97	0.90	0.90	0.97	0.97
Gardettos Snkns	0.92	0.88	0.88	0.93	0.93
Snackwell	0.93	0.87	0.87	0.93	0.93
Payday	0.78	0.77	0.77	0.78	0.78
Peter Pan (Crck)	0.80	0.81	0.81	0.80	0.80
Oreo	0.89	0.84	0.84	0.89	0.89
M&M	1.00	0.91	0.91	1.00	1.00
Crunch Nestle	0.96	0.87	0.87	0.96	0.96
Lays Potato Chip	0.95	0.88	0.88	0.95	0.95
Ding Dong	1.07	0.92	0.92	1.07	1.07
Banana Nut	1.05	0.92	0.92	1.06	1.06
Choc Donuts	1.08	0.92	0.92	1.08	1.08
Kar Nut Sweet & Salty	1.04	0.93	0.93	1.04	1.04
Frito Corn Chip	0.96	0.89	0.89	0.96	0.96
Caramel Crunch	0.95	0.88	0.88	0.95	0.95
Starbursts	1.02	0.93	0.93	1.02	1.02
Gma Mini cookie	1.21	0.81	0.81	1.21	1.21
Munchies	1.00	0.91	0.91	1.00	1.00
Lays Baked Potato	1.09	0.89	0.89	1.10	1.10
Skittles	1.03	0.94	0.94	1.03	1.03
Frito Jalepeno	0.89	0.85	0.85	0.89	0.89
Gma Choc Chip	1.05	0.94	0.94	1.06	1.06

Table 7: Estimates of d_j's

Best Substitute For	Logit	EM Nested Logit
Dorito Nacho	Snickers	Rold Gold Pretzels
Frito Corn Chips	Snickers	Rold Gold Pretzels
Grandmas Choc Chip	Snickers	Grandma's Oatmeal Raisin
Twix	Snickers	Snickers
M&M Peanut	Snickers	Snickers
Hostess Donuts	Snickers	Strawberry Poptart

Table 8: Substitution Patterns

Best Substitute For	BLP	EM-BLP
Dorito Nacho	Cheeto Crunchy	Dorito Guacamole
Frito Corn Chips	Cheeto Crunchy	Cheeto Crunchy
Grandmas Choc Chip	Grandma's Choc	Grandma's Choc
Twix	Snickers	Snickers
M&M Peanut	Snickers	M&M Plain
Hostess Donuts	Hostess Ban Nut Muffin	Hostess Ding Dong

Table 9: Substitution Patterns (cont.)

Category	Product	SO %	Observed Sales		Predicted Sales	
			All	Full Only	Full Avail	Adjusted
Snack	Gma Choc Chip	0%	6.7	6.8	6.7	6.7
	Chips Ahoy Ss	0.63%	7.7	6.5	7.6	7.7
	Knotts Raspberry	0.85%	5.1	5.5	5.1	5.4
	Gma Oatmeal Raisin	0%	4	4.8	4.1	4.6
	Nuter Butter Bites	2.3%	2.2	2.4	2	2.3
Grab	BG Dorito Nacho	5.55%	3.6	6	3.7	6.2
	Sunchip Harvest	7.54%	6.3	5.1	6.2	7.1
	Rold Gold	1.42%	5.1	5.1	5.1	5.1
	Cheeto Crunchy	2.33%	4.7	4.9	4.7	4.6
	Dorito Guacamole	2.33%	4.1	4.4	4.1	4
	Ruffles Cheddar	11.6%	4	4.4	4	4.9
	Munchies	0%	5	4.1	4.9	4.8
	Other Grab		21	23.2	22	19.7
Candy (1)	Snickers	0.42%	7.9	10.8	8.1	8.2
	Twix Carmel	0.42%	10.1	9.9	10	10
	M&M Peanuts	0%	6.3	5.3	6.2	6.2
	Big Kit Kat	0%	3.6	4.2	3.6	3.7
	M&M	0%	3.2	3.7	3.2	3.2
	Other Candy (1)		8	7.2	7.9	8
Candy (2)	Peanuts	0%	4.3	4.7	4.2	4.3
	Skittles	0.85%	5.3	4	5.2	5.3
	Kar Nut Sweet & Salty	7.58%	4	3.3	4	4.3
	Starbursts	0%	2.6	3.1	2.6	2.7
	Oreo Cookie	0%	3.6	3.5	3.5	3.5
Pastry	Donuts Chocolate	15.08%	3.6	5.2	3.5	6.3
	Hostess Pastry	14.84%	4.7	5	4.6	5.4
	Ding Dong	0%	4.5	4.6	4.5	4.5
	Banana Nut	2.54%	3.8	4	3.8	4
	Poptart	5.59%	3.2	4	3.2	4.2

Table 10: Observed and Predicted Average Weekly Sales for Representative Machine

	Regime A -PS	Regime B-PS
Full Availability	0.32	0.46
Ignore Missing Data	0.64	0.81
EM Adjusted	0.69	0.98

Table 11: Expected Hourly Profit Loss of Stockouts

	Logit	Nested Logit	EM Nested	BLP	EM BLP
1	0.02	0.02	0.03	0.03	0.02
100	3.41	2.82	3.11	3.29	3.27
500	17.12	18.21	19.33	19.67	20.43
1000	36.15	43.23	49.44	44.23	50.21

Table 12: Losses from Typical Stockout Patterns

	MSE	Adj R^2	# Parameters
Prod	0.615	0.989	50
Prod, DoW	0.612	0.98	350
Prod, ToD	0.611	0.974	298
Prod,Dow,Tod	0.606	0.959	2044
Mach,Prod	0.606	0.959	1978
Mach,Prod,DoW	0.598	0.934	11692
Mach,Prod,ToD	0.592	0.915	13645
Mach,Prod,Dow,Tod	0.566	0.824	74420

Table 13: Heterogeneity of q_{jt}

	MSE	Adj R^2	# Parameters
DoW	77.211	0.996	7
ToD	77.257	0.997	6
Dow,ToD	77.034	0.991	42
Machine	73.138	0.893	50
Mach DoW	72.364	0.873	350
Mach,ToD	72.805	0.884	300
Mach,Dow,ToD	71.019	0.827	2100

Table 14: Heterogeneity of M_t

References

AGUIRREGABIRIA, V. (1999): "The Dynamics of Markups and Inventories in Reatailing Firms," *Review of Eocnomic Studies*, 66, 278–308.

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 (2006): "Discrete Choice Models with Multiple Unobserved Choice Characteristics," .

BERRY, S. (1992): "An Esitmation of a Model of Entry in the Airline Industry," *Econometrica*, (60), 889–917.

——— (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25(2), 242–261.

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.

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.

DRAGANSKA, M., AND D. JAIN (2004): "A Likelihood Appraoch to Estimating Market Equilibrium Models," *Management Science*, 50(5), 605–616.

HADLEY, G., AND T. WHITMAN (1963): *Analysis of Inventory Systems*. Prentice Hall.

HANSEN, L. P. (1982): "Large Sample Propertes of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HARTLEY, H. (1958): "Maximum Likelihood from Incomplete Data," *Biometrics*, 14(2), 174–194.

Liu, C., and D. B. Rubin (1994): "The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence," *Biometrika*, 81, 633–648.

Matzkin, R. (1992): "Nonparametric and Distribution-Free Estimation of Binary Threshold Crossing Models," *Econometrica*, 60(2), 239–270.

McFadden, D., and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470.

Musalem, A., E. T. Bradlow, and J. S. Raju (2006): "Bayesian Estimation of Random-Coefficients Chocie Models Using Aggregate Data," .

Nevo, A. (2000a): "Mergers with differentiated products: the case of the ready-to-eat cereal industry," *RAND Journal of Economics*, 31(3), 395–421.

——— (2000b): "A Practitioner's Guide to Estimation of Random Coefficients Logit Models of Demand (including Appendix)," *Journal of Economics and Management Strategy*, 9(4), 513–548.

——— (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342.

Nevo, A., and I. Hendel (2007a): "Measuring Implications of Sales and Consumer Inventory Behavior," *Econometrica*.

——— (2007b): "Sales and Consumer Inventory," *The RAND Journal of Economics*.

Nevo, A., and C. Wolfram (2002): "Why Do Manufacturers Issue Coupons? An Empirical Analysis of Breakfast Cereals," *The RAND Journal of Economics*, 33(2), 319–339.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Rubin, D. B. (1976): "Inference and Missing Data," *Biometrika*, 63, 581–592.