Demand Estimation Under Incomplete Product Availability[*]

Christopher T. Conlon
Julie Holland Mortimer

October 4, 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, J.P. Dube, 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. [Anticipate the fact that the method generalizes too.]

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

Let y_i denote the purchase of consumer i, and let x_i denote the relevant observables. Generally we observe some data (y_i, x_i). Typically we think of y_i being categorical where it takes on one of several discrete values $j = (0, 1, \dots, J)$. In an abuse of notation we also let $y_{ij} = 1$ for the chosen product and $y_{ik} = 0$ for the product which was not chosen. Thus y is both an index for which product was chosen as well as an indicator for the chosen product. This leads to the first assumption:

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

[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).

For simplicity, we'll also denote the set of possible choices as J, the powerset of $\{0, 1 \ldots, J\}$ as A, and $a \in A$ as the set of available products. Continuing the standard abuse of notation, we'll also consider a to be a $J \times 1$ vector that takes on value 1 in the jth position if the product is available and zero otherwise.

Without making any parametric assumptions we can write down a multinomial likelihood of seeing an individual choose product j as: [2]

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

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

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

We could begin to think about doing estimation here, but it would uninformative. It is easy to see that the likelihood would be maximized by setting $p_{ij} = 1$ if individual i chose good j and zero otherwise. Unfortunately this tells us nothing about goods that were not chosen. On the other hand if we assume all of the choices were made by the same consumer, then the maximum likelihood estimator for $\hat{p}_{ij} = \frac{1}{n} \sum_{i=1}^{n} y_{ij}$, which is simply the empirical frequency.

We could think about a compromise between the two extremes where we allowed choice probabilities $p_{ij} = p_j(\tilde{x}_i, \theta)$.[3] In other words we let an individual's probability of choosing product j depend on some observable $\tilde{x}_i$ as well as the unknown parameters. The $\tilde{x}_i$ might include information about the individual consumer (z_i), information about the available products (a_i), and perhaps information about the situation under which the choice took place (x_i).[4] In the IO literature we typically parameterize p so that we can do out of sample

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

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

[4]Often times we choose $p(\cdot)$ to be a function that is linear in the $\tilde{x}_i$'s, because this mapping takes us from a linear model in $\mathbb{R}^n$ to a categorical choice model, this is called a Generalized Linear Model (GLM), and $p(\cdot)$ is typically referred to as a link function.

prediction exercises that are of interest or to avoid the "curse of dimensionality".[5]

3.1 Exchangeability

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

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

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

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

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

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

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

An example that violates this might be if we believe that consumers had pent-up demand following a stockout. In this case their choice probability for a good would be higher in periods following a stockout. Unless this was captured in the x, then the likelihood function would depend on the order in which the y_i were observed.

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

[5] A common form for $p(\cdot)$ is the logit form: $p_j(\tilde{x}_i, \theta) = \frac{\exp(\delta_j \beta)}{1 + \sum_{k \in a_i} \exp(\delta_k \beta)}$.

Exchangeability requires that the population of consumer preferences we're sampling from cannot change within a period of observation, denoted as t. Without this assumption, we cannot tell apart stock-out events from atypical consumers. 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 between 4pm and 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.[6]

3.2 Aggregation

Often times we'll observe $(y_{jt}, \tilde{x}_t) = (\sum_{i \in t} y_{it}, \tilde{x}_{it} = \tilde{x}_t)$. In other words, we observe only aggregate data for some period t.[7] It is convenient to treat $\mathbf{y_t}$ as a stacked vector of y_{jt}'s. In general, if we believe that within a period t, consumers are *exchangeable* then we can consider aggregate data without loss. In order to do this we must have that $\tilde{x}_{it} = \tilde{x}_t$ for $i \in t$, or that within our level of aggregation the observables are fixed. There is nothing thus far that requires our periods be contiguous in geography or time, only that they have the same observable $\tilde{x}_t$.[8]

In the case of the logit, this is the same as assuming that for all consumers $\forall i \in t \; \Phi_i = \Phi_t$ or $(a_i, x_i) = (a_t, x_t)$. Thus so long as the observables and choice set are constant within a period, we can consider aggregate data is if it were individual purchase data without loss and write this log-likelihood as:

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

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

[6] 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, for which short-term heterogeneity in the population gets "averaged out" in the overall distribution of consumer preferences.

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

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

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

The key implication of assumptions 1 and 2 is that p_{jt} is constant within a period and does not depend on the realizations of other consumers' choices y_{ijt}. Another immediate implication is that we can reorder the unobserved purchase decisions of individual consumers within a period t. Now, we apply assumption 2 again 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 the set of available products a_t, and the exogenous variables x_t.*

Taking assumptions 1, 2 and 3, we have the following result.

Theorem 1.

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

In words, we observe data (y_t, x_t, a_t) aggregated to some level. As long as (x_t, a_t) are fixed in our level of aggregation, we could act as if we had data on independent consumers. Then we can reconstruct a new level of aggregation t' where we aggregate over all observations with the same (x_t, a_t) pairs without changing the value of the likelihood. This gives us some new "pseudo-(period) observations". Typically we say that $y_{jt'}$ (which is just the sum over all y_{jt} with the same (a, x) values) is a sufficient statistic for the likelihood since it contains all of the information about y_i that we need to evaluate the likelihood function.

From now on, we'll assume that any dataset $(\mathbf{y_i}, \tilde{\mathbf{x}}_\mathbf{i})$ can be reconstructed in this minimal sufficient statistic way, and that those groupings (not the original aggregation in the data) are subscripted by t rather than t'. In fact we'll assume that we have a function S (defined by Theorem 1) which takes the vector of observed periods $(\mathbf{y_t}, \mathbf{x_t}, \mathbf{a_t})$ and converts them into their minimal sufficient statistic representation $(\mathbf{y_{t'}}, \mathbf{x_{t'}}, \mathbf{a_{t'}})$.

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

This leads to the following corollary:

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

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

4 Adjusting for Stockouts

4.1 Perpetual Inventory/Observed Choice Set Heterogeneity

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

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

$$
\begin{aligned}
l(\mathbf{y}|\theta,\mathbf{M},\mathbf{a},\mathbf{x}) &\propto \sum_{\forall t}\left(\sum_{\forall j\in a_s}\ln p_j(\theta,a_s,x_t)\sum_{\forall i:(a_{it},x_{it})=(a_s,x_t)} y_{ji}+\sum_{\forall j\in a_t}\ln p_j(\theta,a_t,x)\sum_{\forall i:(a_{it},x_{it})=(a_t,x_t)} y_{ji}\right)\\
l(\mathbf{y}|\theta,\mathbf{M},\mathbf{a},\mathbf{x}) &\propto \sum_{\forall t}\left(\sum_{\forall(a,x)}\sum_{\forall j\in a}\ln p_j(\theta,a,x)\sum_{\forall i:(a_i,x_i)=(a,x)} y_{ji}\right)\\
&= \sum_{\forall t}\left(\sum_{\forall(a,x)}\sum_{\forall j\in a} y_{j,(a,x)}\ln p_j(\theta,a,x)\right) \qquad (3)
\end{aligned}
$$

4.2 Periodic Inventory/ Latent Choice Set Heterogeneity

As has already been discussed, many retail environments observe inventories periodically. This presents additional challenges when investigating stock-out events, 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. As in the perpetual case, we could denote the set of available choices at the beginning of period t by a_s, and the set remaining at the end of t by a_t. What we'd like to do is assign the sales in period t to each (a_s, a_t) regime, but because inventory is observed only periodically, we can't.

Recall that the likelihood tells us the probability of y conditional on what we observed (a, x) for a guess of the parameters θ

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

But if $\mathbf{a}$ is not fully observed, we can't condition it out. The standard thing to do is to partition $\mathbf{a}$ into observable and unobservable pieces, that is for each individual observation i in the dataset, a_i is either known or it isn't. We can then divide the entire dataset into two subsets T_{mis} and T_{obs} and either $i \in T_{obs}$ or $i \in T_{mis}$. The convention is to define $(\mathbf{y_{obs}}, \mathbf{x_{obs}}, \mathbf{a_{obs}})$ as the vector of fully observed data (ie: $\forall i \in T_{obs}$) and $(\mathbf{y_{mis}}, \mathbf{x_{mis}}, \mathbf{a_{mis}})$ as its complement (ie: $\forall i \in T_{mis}$). It should be clear that we're using T_{obs} and T_{mis} because we can easily do the same thing for aggregate data t instead of just individual data without any additional difficulty.[9]

The likelihood now requires that we integrate out uncertainty regarding the missing data, since we want to know the likelihood of the the independent variable $\mathbf{y}$ conditional on the observed data $(\mathbf{x}, \mathbf{a_{obs}})$ for a given parameter value θ:

$$\begin{aligned} l(\mathbf{y}, \mathbf{a}|\mathbf{a_{obs}}, \mathbf{x}, \theta) &= l(\mathbf{y}|\mathbf{a}, \mathbf{x}, \theta) g(\mathbf{a}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \\ l(\mathbf{y}|\mathbf{a_{obs}}, \mathbf{x}, \theta) &= \int l(\mathbf{y}, \mathbf{a}|\mathbf{a_{obs}}, \mathbf{x}, \theta) \partial \mathbf{a} \end{aligned}$$

The problem is that evaluating this integral is difficult. For one, it involves the joint distribution of the a_i's $g(\mathbf{a}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta)$ which we haven't specified. The other is that this is a vector integral and the dimension is large. Technically the dimension is n, the number of observations in the dataset.

$$\begin{aligned} l(\mathbf{y}|a_{obs}, \mathbf{x}, \theta) &= l(\mathbf{y}|\mathbf{a}, \mathbf{x}, \theta) g(\mathbf{a}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \\ &= \int \left[\sum_{i=1}^{n} l(y_i|a_i, x_i, \theta) g_i(a_i|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \right] \partial \mathbf{a} \\ &= \sum_{i=1}^{n} \left[\int l(y_i|a_i, x_i, \theta) g_i(a_i|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \partial \mathbf{a} \right] \\ &= \sum_{i \in T_{obs}} l(y_i|a_i, x_i, \theta) + \sum_{i \in T_{mis}} \left[\int l(y_i|a_i, x_i, \theta) g_i(a_i|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \partial \mathbf{a} \right] \\ &= l(\mathbf{y_{obs}}|\mathbf{a_{obs}}, \mathbf{x_{obs}}, \theta) + \int l(\mathbf{y_{mis}}|\mathbf{a_{mis}}, \mathbf{x_{mis}}, \theta) g(\mathbf{a_{mis}}|\mathbf{a_{obs}}, \mathbf{y}, \mathbf{x}, \theta) \partial \mathbf{a_{mis}} \end{aligned}$$

It is easy to see that we can reduce the dimension of the integral to only the number of missing observations, $n_{mis} = \dim(T_{mis})$ without too much difficulty. (Since the probability of observed choice sets is 1 if we observe them and zero otherwise). However, this is still a high dimensional integral that we must evaluate each time we evaluate the likelihood at a new parameter value θ. Moreover it relies on a distribution $g(\cdot)$ that we don't know much about.

One solution is to parameterize $g(\cdot)$ in some convenient ad-hoc way, to make the integral easier. We might assume that the missing values are (exchangeable) independent conditional on the observable $(\mathbf{y}, \mathbf{x})$ and the parameters θ, in other words, $g(\mathbf{a}|a_{obs}, \mathbf{y}, \mathbf{x}, \theta) = \prod_{i=1}^{n} g(a_i|\mathbf{y}, \mathbf{x}, \theta)$. This allows us to consider a series of n_{mis} single dimensional integrals

[9] The other technical point here that should be made clear is that the only thing that is ever unobserved is the value of a_i, we assume we still perfectly observe (y_i, x_i) just like we did before.

instead of one large n_{mis} dimensional joint integral.

$$l(\mathbf{y}|a_{obs},\mathbf{x},\theta) = \sum_{i\in T_{obs}} l(y_i|a_i,x_i,\theta) + \sum_{i\in T_{mis}} \left[\int l(y_i|a_i,x_i,\theta) g(a_i|\mathbf{y},\mathbf{x},\theta)\partial a_i\right]$$

4.3 The Case of Stockouts

Stockouts represent an important special case of unobservable choice set heterogeneity. What we haven't done thus far is specified a distribution for $g(a|\mathbf{y},\mathbf{x},\theta)$. What's interesting about the stockout case is that rather than making an ad-hoc assumption to simplify the integral, the model will specify the distribution.

The question we want to answer is "What proportion of consumers faced choice set a_l out of possible choice sets A_t in a given aggregate observation?". In the stockout case, the change in choice set is endogenous, that is, happens when y_{jt} exceeds its capacity ω_{jt}. If we limit ourselves to a single stockout for now, and there are M_t consumers in a particular period of observation, then we can ask, how many of the M_t consumers saw choice set a and how many saw choice set a'. Another way to think about this is to ask, how many consumers did it take to sell ω_{jt} units of product j under availability set a?

This is exactly the definition of the negative binomial (number of failures k until r successes are observed). In the stockout case, we know that the stockout happened before M_t consumers arrived at the machine. In other words we have a negative binomial, conditional on the fact that $k + r \leq M_t$.[10]

The definition of the negative binomial is:

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

And the conditional negative binomial:

$$\begin{aligned} Pr(z = k|z \leq K) &\sim \frac{NegBin(r, p)}{NegBinCDF(K, r, p)} \\ h(k', r, p) &= \frac{f(k', r, p)}{\sum_{k=1}^{K} f(k, r, p)} \end{aligned}$$

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

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

$$\begin{aligned} \lambda_t &= \frac{k+r}{M_t} \\ Pr\left(\lambda = \frac{k+r}{M_t} | \lambda \leq 1\right) &\sim \frac{NegBin(k, y_{jt}, p_j(a, x_t, \theta))}{NegBinCDF(M, y_{jt}, p_j(a, x_t, \theta))} \end{aligned}$$

This is the exact distribution for the fraction of consumers facing a stockout. In other words this is $g(\lambda|A_t, \mathbf{y}, \mathbf{x}, \theta)$. To evaluate integrals over this, we will need to evaluate the density at all M_t points, as long as M_t is reasonably small this is easy. Since the number of vending machine customers is generally small over short periods of time, this is not a problem. If it were a problem, we could think about approximating this distribution with Monte Carlo draws, or a finite number of gridpoints. Another possibility is that as M gets large, continuous approximations (rather than an exhaustive evaluation of every discrete point) become more accurate and are generally easier to compute.

Now, we hope to finally be able to evaluate the mixture likelihood (with availability a before the stockout and a' after).

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

Evaluating this is not too difficult, since it is just a sum over j and a sum over M_t. The distribution $g(\cdot)$ depends only on data from its own observation t, and doesn't depend on realizations of a_t from other observations. This is an implication of the model for demand under stockouts, not an assumption. Thus, when analyzing stock-out events, the distribution across availability regimes is already directly implied by the model.

4.4 Augmented Model

Denote the possible choice sets in period t as a_l, and recall the augmented model from the previous section when we observe the sales under each choice set y_{jta_l}.

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

In this case the observed sales for observation t are distributed multinomially (across $J \times \dim(A_t)$ cells). In fact we could consider breaking up the data into $\dim A_t$ observations each with J cells, and then the likelihood takes on exactly the same form as the case without any unobservable choice set heterogeneity. (Restatement of Theorem 1).

The problem with periodic data is that y_{jta_l} is not actually observed, so we cannot evaluate $l_t(y_{ta_l}, A_t, x_t, \theta)$, but we can replace it with a consistent estimator (its expectation). Then the likelihood becomes:

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

This is now a likelihood function of only the observables (for a fixed value of θ'). By iterating back and forth between computing this expectation at some value of θ', plugging this in and then maximizing the complete-data likelihood over θ, and updating θ' we can obtain consistent parameter estimates for θ. This is the well known E-M Algorithm of (Dempster Laird and Rubin).

It should be clear that this approach is no different from integrating out λ above, except that evaluating this expectation is much easier.

We want to evaluate:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

This is exactly the differentiated products structure found in many IO models (Berry 1994, Berry, Levinsohn, and Pakes 1995). These models have some very nice properties. The first is that any RUM can be approximated arbitrarily well by this "logit" form (McFadden and Train 2000). This also means that the logit ($\mu_{ijt} = 0$) and 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 (2000) 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:

$$\begin{aligned} \delta_{jt} &= \underbrace{X_j\beta + \xi_j} + \xi_t + \Delta\xi_{jt} \\ &= \quad d_j \quad + \xi_t + \Delta\xi_{jt} \end{aligned} \tag{4}$$

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 (2000) 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 that 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.[11] 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 Maximum Simulated Likelihood Estimator (MSLE)

For the multinomial choice model the MSLE is easy to define. Begin with some random or quasi-random normal draws v_{ik} for each t in the dataset. For a given θ we can compute the

[11] 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 $\Delta\xi_{jt}$.

average choice probability across draws.

$$\begin{aligned} l(\theta) &= \sum_t \sum_j y_{jt} \ln \hat{p}_{jt}(\theta) \\ \hat{p}_{jt} &= \frac{1}{ns} \sum_{i=1}^{ns} p_{ijt}(x, \theta, \nu_i) \\ p_{ijt} &= \frac{\exp[d_j + \sum_l \sigma_l v_{il} x_{jl}]}{1 + \sum_{j \in a_t} \exp[d_j + \sum_l \sigma_l v_{il} x_{jl}]} = \frac{z_j(d_j)\mu_{ij}(\sigma)}{1 + \sum_j z_j(d_j)\mu_{ij}(\sigma)} = \frac{z_j(d_j)\mu_{ij}(\sigma)}{D} \end{aligned}$$

Additional details are provided in the appendix.

5.3 Heterogeneity

Thus far, we've done everything conditional on x_t. In one sense, this is useful to show that our result holds for the case of conditional likelihood, but it is also of practical significance to our applied problem. Since periods in retail datasets may be short, it is likely that choice probabilities may vary substantially over periods. Over long periods of time (such as annual aggregate data) these variations get averaged out. The distribution of tastes over a long period is essentially the combination of many short-term taste distributions, and this is often the basis of estimation (ie., in the case of limited data we would estimate the long run distribution). With high frequency data we are no longer so limited and can address this additional heterogeneity by conditioning on x_t. [Note: discussion of sampling error should be addressed here again.] We might think that x_t includes information such as the time of day, day of the week, or local market identifiers. Depending on how finely data are observed, not accounting for this additional heterogeneity may place a priori unreasonable restrictions on the data.

We can model 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 local market 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 2000), 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 effect on the size of the market. Parameterizing M has a long history in the literature (Berry 1992). In practice, it is 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.

5.4 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).[12]

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. [Note: we can get price coefficients from price variation, as usual.] 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,

[12] 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.

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 variation 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.

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.[13]

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 the spring semester of 2003 through the fall semester of 2004. [Confirm Fall 2003.]

[13] 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.

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 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 [14] 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 Table 1.

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,[15] 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 (2000) for RTE cereal.

The retail prices observed in the vending machine are constant over time and across

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

[15] 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.

broad groups of products as shown in Table 1. 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 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 Table 1. By examining Table 1, 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™chips (about 70%). The product with the highest markup is the Peter Pan crackers, which have an average markup of nearly 82%.

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 5 reports the results of a regression of stockout rates on starting inventory levels. We report results for Probit and OLS with and without product and machine fixed effects. We find that an additional unit of inventory at the beginning of a service period reduces the chance of a stockout in that product by about 1%. A full column of candy bars usually contains 20 units. This means that the OLS (fixed effects) probability of witnessing a stockout from a full machine in 3-day period is $.311 - .0113 * 20 = 8.5\%$. For a machine with a starting inventory of five units, the predicted chance of a stockout is one in four.

In table 4, we compute the average 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 4-hourly cost to be about \$0.53 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 3. These estimate the costs per stockout at around \$0.92 per chocolate candy bar to \$2.38 per non-chocolate candy bar. Column 2 examines the effect of a stockout in the category with the most stockouts and estimates this effect to be about \$1.40 per 4-hour period. Column 4 also includes 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 product characteristics in the nonlinear part. These are: calories, fat, carbs, sugar, salt, chocolate and cheese.

Tables 8 and 9 reports the corrected values of many of the product dummies (the d_j's) under each of the different models. (More on this to come....)

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. (More to come...)

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.

Product	Category	% SO (Lower)	% SO (Upper)	p	c	Share	Avg D Sales	No. Mach.
PopTart	Pastry	4.76	5.62	1.00	0.35	3.65	0.74	59
Choc Donuts	Pastry	14.42	16.32	1.00	0.46	2.93	0.67	58
Ding Dong	Pastry	13.27	15.09	1.00	0.46	2.76	0.60	57
Banana Nut Muffin	Pastry	8.95	10.10	1.00	0.40	2.68	0.54	59
Rice Krispies	Pastry	2.18	2.45	1.00	0.31	1.95	0.40	59
Pastry	Pastry	9.69	11.31	1.00	0.46	0.78	0.87	24
Gma Oatmeal Raisin	Cookie	3.19	3.60	0.75	0.23	2.75	0.58	57
Chips Ahoy	Cookie	1.55	1.81	0.75	0.25	2.50	0.51	58
Knotts Raspberry	Cookie	0.96	1.10	0.75	0.19	1.82	0.38	57
Nutter Butter Bites	Cookie	0.95	1.02	0.75	0.26	1.79	0.38	56
Gma Choc Chip	Cookie	0.96	1.05	0.75	0.22	1.38	0.68	57
Gma Mini Cookie	Cookie	1.75	2.05	0.75	0.21	0.87	0.56	54
Gma Car Ch Chip	Cookie	2.09	2.49	0.75	0.23	0.67	0.58	54
Snickers	Candy (C)	1.00	1.22	0.75	0.33	8.29	1.66	59
Twix	Candy (C)	1.18	1.34	0.75	0.33	6.08	1.22	59
M&M Peanut	Candy (C)	2.20	2.59	0.75	0.33	4.68	0.94	59
Reese's Cup	Candy (C)	1.33	1.46	0.75	0.33	2.36	0.47	59
Caramel Crunch	Candy (C)	1.60	1.73	0.75	0.33	2.21	0.44	58
Kit Kat	Candy (C)	1.02	1.15	0.75	0.33	2.18	0.44	59
Hershey Almond	Candy (C)	1.21	1.31	0.75	0.33	1.77	0.36	59
M&M	Candy (C)	1.04	1.12	0.75	0.33	1.71	0.52	59
Babyruth	Candy (C)	2.07	2.35	0.75	0.28	0.57	0.34	56
Kar Nut Sweet/Salt	Candy (NC)	2.51	2.81	0.75	0.22	2.99	0.60	59
Starburst	Candy (NC)	0.54	0.71	0.75	0.33	2.95	0.79	59
Snackwell	Candy (NC)	1.03	1.14	0.75	0.28	1.70	0.35	59
Skittles	Candy (NC)	1.05	1.27	0.75	0.34	1.44	0.70	59
Payday	Candy (NC)	0.14	0.14	0.75	0.33	1.25	0.43	57
Oreo	Candy (NC)	0.52	0.57	0.75	0.22	1.06	0.22	59
Peter Pan (Crck)	Candy (NC)	0.30	0.31	0.75	0.12	0.91	0.35	52
Peanuts	Candy (NC)	0.77	0.83	0.75	0.26	0.83	0.42	56
Hot Tamales	Candy (NC)	2.21	2.60	0.75	0.27	0.52	0.42	56
Rold Gold	Chips	6.86	8.11	0.90	0.27	3.97	0.80	59
Sunchip Harvest	Chips	7.08	8.25	0.90	0.27	3.81	0.77	59
Gardetto Snackens	Chips	0.80	0.85	0.75	0.26	3.41	0.95	58
Dorito Nacho	Chips	2.80	3.23	0.90	0.27	3.39	0.68	59
Cheeto Crunchy	Chips	6.47	7.37	0.90	0.27	3.32	0.67	59
Ruffles Cheddar	Chips	2.21	2.59	0.90	0.27	2.67	0.55	59
Fritos	Chips	4.07	4.53	0.90	0.27	1.93	0.39	59
Lays Potato Chip	Chips	4.46	4.96	0.90	0.19	1.72	0.35	58
Misc Chips 2	Chips	2.45	2.74	0.90	0.28	1.35	0.29	59
Munchies Hot	Chips	1.64	1.82	0.75	0.25	1.20	0.76	52
Munchies	Chips	2.82	3.19	0.90	0.25	1.18	0.43	57
Misc Chips 1	Chips	2.06	2.25	0.90	0.27	1.16	0.49	59
Dorito Guacamole	Chips	1.22	1.38	0.90	0.28	0.86	0.42	58

Table 1: Summary of Products and Markups

	(1)	(2)	(3)	(4)
SO at end	-0.441		-0.606	
	(0.35)		(0.40)	
SO at beginning		-0.387		-0.538
		(0.30)		(0.35)
Cookie			5.990	5.926
			(4.12)	(4.11)
Chips			7.440**	7.386**
			(3.39)	(3.38)
Candy (C)			-7.122*	-7.202*
			(4.07)	(4.07)
Candy (NC)			0.785	0.714
			(4.19)	(4.18)
Constant	66.08***	66.08***	64.72***	64.80***
	(1.53)	(1.53)	(4.13)	(4.12)
Observations	44	44	44	44
R^2	0.0369	0.0372	0.5595	0.5604

Table 2: Regression of Markup on Percent time Stocked Out

	(1)	(2)	(3)	(4)
# Products Stocked Out	-0.528***	-0.238***	-0.228***	-0.364***
	(0.0076)	(0.011)	(0.012)	(0.012)
# SOs in Cat. w/ Max		-1.396***		
		(0.035)		
# SO, pastry			-1.999***	-4.531***
			(0.045)	(0.070)
# SO, cookie			-1.595***	-3.939***
			(0.098)	(0.11)
# SO, chips			-1.066***	-3.765***
			(0.041)	(0.066)
# SO, Candy (C)			-0.917***	-3.140***
			(0.074)	(0.086)
# SO, Candy (NC)			-2.381***	-3.876***
			(0.13)	(0.13)
Max. Cat. SO = 2				4.142***
				(0.14)
Max. Cat. SO = 3				9.054***
				(0.23)
Max. Cat. SO >3				17.51***
				(0.32)
Constant	6.307***	6.599***	6.671***	7.533***
	(0.033)	(0.034)	(0.034)	(0.037)
Observations	181103	181103	181103	181103
R^2	0.0256	0.0341	0.0368	0.0529

Table 3: Regression of Profit on Stock-Out Variables

	(1)	(2)	(3)	(4)
# Products Stocked Out	-0.560***	-0.282***	-0.269***	-0.398***
	(0.0078)	(0.010)	(0.012)	(0.012)
# SOs in Cat. w/ Max		-1.332***		
		(0.034)		
# SO, pastry			-1.960***	-4.386***
			(0.044)	(0.068)
# SO, cookie			-1.631***	-3.875***
			(0.095)	(0.10)
# SO, chips			-0.989***	-3.567***
			(0.039)	(0.064)
# SO, Candy (C)			-0.864***	-2.995***
			(0.072)	(0.083)
# SO, Candy (NC)			-2.353***	-3.773***
			(0.12)	(0.13)
Max. Cat. SO = 2				4.005***
				(0.13)
Max. Cat. SO = 3				8.659***
				(0.22)
Max. Cat. SO >3				16.69***
				(0.31)
Observations	181103	181103	181103	181103
R^2	0.2294	0.2359	0.2384	0.2506

Table 4: Regression of Profit on Stock-Out Variables, Including Machine FE

dy/dx	OLS	OLS	OLS	Probit	Probit	Probit
Inventory	-0.0167***	-0.0087***	-0.0113***	-0.0143***	-0.0072***	-0.0097***
(SE)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0004)	(0.0004)
Elapsed time	0.0020***	0.0022***	0.0028***	0.0017***	0.0018***	0.0023***
(SE)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	0.3390***	0.2500***	0.3110***	-0.2700	-0.5690	-0.2170
(SE)	(0.0038)	(0.0085)	(0.0120)	(0.0168)	(0.0438)	(0.0632)
Observations	98900	98900	98900	98900	98900	98900
Fixed Effects	-	Prod	Prod-Mach		Prod	Prod-Mach
R-squared	0.049	0.133	0.306	0.059	0.156	0.179

Table 5: Table of Stockout Rates on Starting Inventory

	Full	Ignore	SO-LB	SO-UB	EM
σ's					
Pastry	0.0000	0.0000	0.2473	0.2601	0.2824
Cookie	0.0421	0.0001	0.0002	0.0002	0.0002
Chips	0.0021	0.0045	0.0137	0.0162	0.0162
Choc.	0.0039	0.0120	0.6563	0.7059	0.6195
Non-Ch.	0.0005	0.0024	0.0815	0.0784	0.1093
No. Choice Sets	362	2911	3345	3862	3923
Log L(.) (*-1000)	2555	2150	2547	2493	2416
Total M (*1000)	3082	2859	3082	3082	3082
Likelihood/M	0.829	0.752	0.827	0.809	0.784

Table 6: Parameter Estimates, Nested Logit

	Full	Ignore	SO-LB	SO-UB	EM
σ's					
Calories	0.0000	0.0005	0.4356	1.0028	0.4871
Fat	0.0003	0.0003	0.0002	0.0002	0.0002
Sodium	0.0000	0.0000	0.0148	0.007	0.0170
Sugar	0.0000	0.0285	0.6003	2.6527	0.7276
Choc	0.0007	0.0012	0.0687	0.0000	0.0753
Cheese	0.0116	0.0174	0.0825	0.0001	0.0905
No. Choice Sets	362	2911	3345	3862	3923
Log L(.) (*-1000)	2555	2150	2547	2493	2416
Total M (*1000)	3082	2859	3082	3082	3082
Likelihood/M	0.829	0.752	0.827	0.809	0.784

Table 7: Parameter Estimates, Random Coefficients

	Full	Ignore	SO-LB	SO-UB	EM	DIFF	
Snickers	1.14	1.11	1.14	1.14	1.14	0.00	
Twix	1.06	1.04	1.07	1.06	1.06	-0.28	
M&M Peanut	1.01	0.99	1.01	1.01	1.01	0.30	
Reeses Cup	0.89	0.87	0.89	0.89	0.88	-0.23	
Caramel Crunch	0.88	0.86	0.88	0.88	0.88	-0.11	
Kit Kat	0.87	0.85	0.87	0.87	0.87	-0.34	
Hershey Almond	0.84	0.83	0.84	0.84	0.84	0.24	
M&M	0.90	0.88	0.90	0.90	0.90	0.00	
Babyruth	0.84	0.82	0.84	0.83	0.84	0.48	
Kar Nut Sweet/Salt	0.93	0.91	0.93	0.93	0.93	-0.11	
Starburst	0.97	0.95	0.97	0.97	0.97	-0.10	
Snackwell	0.84	0.82	0.84	0.84	0.84	-0.12	
Skittles	0.95	0.93	0.95	0.95	0.95	-0.42	
Payday	0.87	0.85	0.87	0.87	0.87	-0.11	
Oreo	0.78	0.76	0.78	0.78	0.78	-0.26	
Peter Pan (Crck)	0.84	0.82	0.84	0.84	0.83	-0.24	
Peanuts	0.86	0.84	0.86	0.86	0.86	-0.46	
Hot Tamales	0.87	0.86	0.88	0.87	0.88	1.04	
Rold Gold	0.98	0.96	0.99	0.98	0.98	0.72	*
Sunchip Harvest	0.97	0.96	0.98	0.97	0.98	0.72	*
Gardetto Snackens	1.01	0.99	1.01	1.01	1.01	-0.20	
Dorito Nacho	0.95	0.93	0.95	0.95	0.95	0.00	
Cheeto Crunchy	0.94	0.93	0.95	0.95	0.95	0.63	*
Ruffles Cheddar	0.91	0.89	0.91	0.91	0.91	0.11	
Fritos	0.86	0.84	0.86	0.86	0.86	0.00	*
Lays Potato Chip	0.84	0.83	0.85	0.84	0.84	0.48	*
Misc Chips 2	0.81	0.83	0.85	0.85	0.85	4.42	
Munchies Hot	0.96	0.87	0.89	0.89	0.89	-7.78	
Munchies	0.86	0.84	0.87	0.86	0.86	-0.12	
Misc Chips 1	0.86	0.85	0.87	0.87	0.87	0.93	
Dorito Guacamole	0.86	0.84	0.87	0.86	0.86	-0.46	
PopTart	0.96	0.95	0.97	0.96	0.97	0.62	*
Choc Donuts	0.95	0.95	0.97	0.96	0.97	2.11	*
Ding Dong	0.93	0.93	0.95	0.93	0.94	1.72	*
Banana Nut Muffin	0.91	0.90	0.92	0.91	0.92	0.88	*
Rice Krispies	0.86	0.84	0.86	0.86	0.86	0.12	
Pastry	0.98	0.97	0.99	0.98	0.99	0.82	*
Gma Oatmeal Raisin	0.92	0.90	0.92	0.92	0.92	0.22	
Chips Ahoy	0.90	0.88	0.90	0.90	0.90	-0.11	
Knotts Raspberry	0.85	0.83	0.85	0.85	0.85	-0.12	
Nutter Butter Bites	0.85	0.83	0.85	0.85	0.85	-0.35	
Gma Choc Chip	0.95	0.92	0.95	0.95	0.94	-0.21	
Gma Mini Cookie	0.92	0.90	0.92	0.91	0.91	-0.44	
Gma Car Ch Chip	0.91	0.90	0.92	0.92	0.92	1.32	

Table 8: d_j's, Nested Logit

	Full	Ignore	SO-LB	SO-UB	EM	DIFF	
Snickers	1.14	1.11	1.14	1.14	1.14	0.00	
Twix	1.06	1.04	1.07	1.06	1.06	-0.09	
M&M Peanut	1.01	0.99	1.01	1.01	1.01	0.20	
Reeses Cup	0.89	0.87	0.89	0.89	0.88	-0.23	
Caramel Crunch	0.88	0.86	0.88	0.88	0.88	0.00	
Kit Kat	0.87	0.85	0.87	0.87	0.87	-0.11	
Hershey Almond	0.84	0.83	0.84	0.84	0.84	0.00	
M&M	0.90	0.88	0.90	0.90	0.90	-0.11	
Babyruth	0.84	0.82	0.84	0.83	0.84	0.24	
Kar Nut Sweet/Salt	0.93	0.91	0.93	0.93	0.93	0.00	
Starburst	0.97	0.95	0.97	0.97	0.97	-0.10	
Snackwell	0.84	0.82	0.84	0.84	0.84	-0.24	
Skittles	0.95	0.93	0.95	0.95	0.95	-0.21	
Payday	0.87	0.85	0.87	0.87	0.87	-0.11	
Oreo	0.78	0.76	0.78	0.78	0.78	0.00	
Peter Pan (Crck)	0.84	0.82	0.84	0.84	0.84	-0.12	
Peanuts	0.86	0.84	0.86	0.86	0.86	-0.23	
Hot Tamales	0.87	0.86	0.88	0.87	0.87	0.92	
Rold Gold	0.98	0.96	0.99	0.98	0.98	0.82	*
Sunchip Harvest	0.97	0.96	0.98	0.97	0.98	0.82	*
Gardetto Snackens	1.01	0.99	1.01	1.01	1.01	-0.20	
Dorito Nacho	0.95	0.93	0.95	0.95	0.95	0.21	
Cheeto Crunchy	0.94	0.93	0.95	0.95	0.95	0.53	*
Ruffles Cheddar	0.91	0.89	0.91	0.91	0.91	0.00	
Fritos	0.86	0.84	0.86	0.86	0.86	0.12	*
Lays Potato Chip	0.84	0.83	0.85	0.84	0.84	0.24	*
Misc Chips 2	0.81	0.83	0.85	0.85	0.85	4.29	
Munchies Hot	0.96	0.87	0.89	0.89	0.89	-7.99	
Munchies	0.86	0.84	0.87	0.86	0.86	-0.23	
Misc Chips 1	0.86	0.85	0.87	0.87	0.87	1.17	
Dorito Guacamole	0.86	0.84	0.87	0.86	0.86	-0.35	
PopTart	0.96	0.95	0.97	0.96	0.97	0.42	*
Choc Donuts	0.95	0.95	0.97	0.96	0.97	2.11	*
Ding Dong	0.93	0.93	0.95	0.93	0.94	1.83	*
Banana Nut Muffin	0.91	0.90	0.92	0.91	0.92	0.77	*
Rice Krispies	0.86	0.84	0.86	0.86	0.86	0.00	
Pastry	0.98	0.97	0.99	0.98	0.99	0.92	*
Gma Oatmeal Raisin	0.92	0.90	0.92	0.92	0.92	0.11	
Chips Ahoy	0.90	0.88	0.90	0.90	0.90	-0.22	
Knotts Raspberry	0.85	0.83	0.85	0.85	0.85	0.00	
Nutter Butter Bites	0.85	0.83	0.85	0.85	0.85	-0.35	
Gma Choc Chip	0.95	0.92	0.95	0.95	0.94	-0.21	
Gma Mini Cookie	0.92	0.90	0.92	0.91	0.91	-0.44	
Gma Car Ch Chip	0.91	0.90	0.92	0.92	0.92	1.32	

Table 9: d_j's, Random Coefficients

	Full	Ignore	SO-LB	SO-UB	EM
Constant	-5.96	-6.14	-6.00	-5.98	-6.02
	(0.24)	(0.23)	(0.23)	(0.23)	(0.23)
Calories	3.55	3.52	3.45	3.40	3.55
	(2.72)	(2.66)	(2.65)	(2.64)	(2.64)
Fat	-3.83	-3.66	-3.59	-3.65	-3.72
	(2.83)	(2.77)	(2.77)	(2.75)	(2.75)
Sodium	-0.10	-0.08	-0.05	-0.10	-0.07
	(0.74)	(0.72)	(0.72)	(0.72)	(0.72)
Sugar	-0.57	-0.38	-0.38	-0.41	-0.42
	(1.28)	(1.25)	(1.25)	(1.24)	(1.24)
Choc	0.20	0.21	0.20	0.20	0.21
	(0.18)	(0.17)	(0.17)	(0.17)	(0.17)
Cheese	0.11	0.17	0.17	0.17	0.17
	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
R^2	0.13	0.16	0.15	0.14	0.16

Table 10: d_j's on Characteristics, Nested Logit

	Full	Ignore	SO-LB	SO-UB	EM
Constant	-5.96	-6.14	-6.00	-5.98	-6.02
	(0.24)	(0.24)	(0.23)	(0.23)	(0.23)
Calories	3.55	3.52	3.45	3.40	3.48
	(2.72)	(2.66)	(2.65)	(2.64)	(2.65)
Fat	-3.83	-3.66	-3.59	-3.65	-3.64
	(2.83)	(2.77)	(2.77)	(2.75)	(2.76)
Sodium	-0.10	-0.08	-0.05	-0.10	-0.06
	(0.74)	(0.72)	(0.72)	(0.72)	(0.72)
Sugar	-0.57	-0.38	-0.38	-0.41	-0.40
	(1.28)	(1.25)	(1.25)	(1.24)	(1.25)
Choc	0.20	0.21	0.20	0.20	0.21
	(0.18)	(0.17)	(0.17)	(0.17)	(0.17)
Cheese	0.11	0.17	0.17	0.17	0.17
	(0.21)	(0.21)	(0.21)	(0.21)	(0.21)
R^2	0.13	0.16	0.15	0.14	0.15

Table 11: d_j's on Characteristics, Random Coefficients

A Estimation Details and the Case of Multiple Stockouts

A.1 Estimation Details

Some preliminary derivatives for the MSLE.

$$\frac{\partial z_j(d_j)}{\partial d_j} = z_j(d_j)$$
$$\frac{\partial \mu_{ij}(\sigma)}{\partial d_k} = \frac{\partial z_j(d_j)}{\partial \sigma} = \frac{\partial z_j(d_j)}{\partial d_k} = 0$$
$$\frac{\partial \mu_{ij}(\sigma)}{\partial \sigma_l} = \mu_{ij}(\sigma)\nu_{il}x_{jl}$$

Now we want the derivatives with respect to d_j and σ_k.

$$\begin{aligned}
\frac{\partial p_{ijt}(\theta)}{\partial d_j} &= \frac{D z_j(d_j)\mu_{ij}(\sigma)}{D^2} - \frac{z_j(d_j)\mu_{ij}(\sigma)z_j(d_j)\mu_{ij}(\sigma)}{D^2} = p_{ijt}(1-p_{ijt}) \\
\frac{\partial l}{\partial d_j} &= \sum_t \sum_j \frac{y_{jt}}{ns} \sum_{i=1}^{ns} \frac{\partial \log p_{ijt}(\theta)}{\partial d_j} \\
&= \sum_t \sum_j \frac{y_{jt}}{ns} \sum_{i=1}^{ns} \frac{1}{p_{ijt}} \frac{\partial p_{ijt}(\theta)}{\partial d_j} = \sum_t \sum_j \frac{y_{jt}}{p_{ijt} \cdot ns} \sum_{i=1}^{ns} p_{ijt}(1-p_{ijt}) \\
&= \sum_t \sum_j y_{jt}(1-\hat{p}_{ijt})
\end{aligned}$$

Additional information coming.

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:

$$
\begin{aligned}
Pr(m_{0,R}, m_{A,R}, m_{AB,R}, R) &= \sum_{x_B=0}^{\omega_B-1} Pr(m_{0,R}|X_A=\omega_A, X_B=x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A}=\omega_B - x_{B,0}) \\
Pr(m_{0,S}, m_{B,S}, m_{AB,S}, S) &= \sum_{x_A=0}^{\omega_A-1} Pr(m_{0,S}|X_B=\omega_B, X_A=x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B}=\omega_A - x_{A,0})
\end{aligned}
$$

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

$$
\begin{aligned}
h(m_0, m_A, m_{AB}, m_B) &= \sum_{x_B=0}^{\omega_B-1} Pr(m_{0,R}|X_A=\omega_A, X_B=x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A}=\omega_B - x_{B,0}) \\
&+ \sum_{x_A=0}^{\omega_A-1} Pr(m_{0,S}|X_B=\omega_B, X_A=x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B}=\omega_A - x_{A,0})
\end{aligned}
$$

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 $m_A, m_B > 0$.

$$
\begin{aligned}
H(m_0, m_A, m_{AB}, m_B) &= \\
& \sum_{m_{0,R}=0}^{M} \sum_{m_{A,R}=0}^{M-m_{0,R}} \sum_{x_B=0}^{\omega_B-1} Pr(m_{0,R}|X_A=\omega_A, X_B=x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A}=\omega_B - x_{B,0}) \\
&+ \sum_{m_{0,S}=0}^{M} \sum_{m_{B,S}=0}^{M-m_{0,S}} \sum_{x_A=0}^{\omega_A-1} Pr(m_{0,S}|X_B=\omega_B, X_A=x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B}=\omega_A - x_{A,0}) \\
= \sum_{m_0=0}^{M} \sum_{m_A=0}^{M-m_0} \sum_{m_B=0}^{M-m_0} \sum_{m_{AB}=0}^{M-m_A-m_B} & \sum_{x_B=0}^{\omega_B-1} Pr(m_{0,R}|X_A=\omega_A, X_B=x_{B,0}) \cdot Pr(m_{A,R}|X_{B,A}=\omega_B - x_{B,0}) \\
&+ \sum_{x_A=0}^{\omega_A-1} Pr(m_{0,S}|X_B=\omega_B, X_A=x_{A,0}) \cdot Pr(m_{B,S}|X_{A,B}=\omega_A - x_{A,0})
\end{aligned}
$$

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.3 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.

The easiest way to look at this is to think about three "goods". The first two are the two products which stock out, which we've labeled a and b. The third good is all other goods. In other words $p_{aog} = p_0 + \sum_{j \in a_t \setminus \{a,b\}} p_j$. Then for case R we have that:

$$Pr(N = m_0 | X_a = \omega_a, X_b = x_{B,0}) = \frac{m_0!}{\omega_a! x_{B,0}! (m_0 - \omega_a - x_{B,0})!} p_a^{\omega_a} p_b^{x_{B,0}} p_{aog}^{m_0 - \omega_a - x_{B,0}}$$

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

We can define the negative multinomial likewise for S. Note that the negative multinomial requires that we fix $x_{B,0}$ the number of units of B sold when both products are available. This, along with ω_B the initial inventory of B gives us the number of units of B available at the beginning of the regime where A has stocked out and B is available which let's us plug directly into the negative binomial formula for the number of consumers in that period. Since both distributions are essentially conditional on $x_{B,0}$ we have to integrate it out which is why we sum them both over $\sum_{x_B=0}^{\omega_B-1}$.

References

Aguirregabiria, V. (1999): "The Dynamics of Markups and Inventories in Retailing 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 Estimation 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.

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

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

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

——— (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.