American Economic Journal: Microeconomics 2013, 5(4): 1–30
http://dx.doi.org/10.1257/mic.5.4.1

Demand Estimation Under Incomplete Product Availability[†]

By Christopher T. Conlon and Julie Holland Mortimer*

Incomplete product availability is an important feature of many markets, and ignoring changes in availability may bias demand estimates. We study a new dataset from a wireless inventory system on vending machines to track product availability every four hours. The data allow us to account for product availability when estimating demand, and provide valuable variation for identifying substitution patterns when products stock out. We develop a procedure that allows for changes in product availability when availability is only observed periodically. We find significant differences in demand estimates: the corrected model predicts significantly larger impacts of stock-out events on profitability. (*JEL* D12, D92, G31, L81, M11)

Incomplete product availability is a common and important feature of markets where products are perishable, seasonal, or have storage costs. For example, retail markets, sporting and concert events, and airlines face important capacity constraints that often lead to stock outs. Not surprisingly, firms in such industries identify inventory management as a critical strategic decision, and consumers cite product availability as a major concern.[1] In these settings, failing to account for product availability not only ignores a useful source of variation for identifying demand parameters, but may also lead to biased estimates of demand. The first source of bias arises from the *censoring* of demand. If a product sells out, the actual demand for a product (at given prices) may be greater than the observed sales, leading to a negative bias in demand estimates. At the same time, during periods of reduced availability of other products, sales of available products may increase. This

* Conlon: Columbia University, Department of Economics, 420 West 118th St., New York, NY 10027 (e-mail: cconlon@columbia.edu); Mortimer: Boston College, Department of Economics, 140 Commonwealth Ave., Chestnut Hill, MA 02467 and National Bureau of Economic Research (e-mail: julie.mortimer.2@bc.edu). We thank Dan Ackerberg, Susan Athey, Steve Berry, Uli Doraszelski, J.P. Dube, Phil Haile, Wes Hartmann, Ken Hendricks, Guido Imbens, Phillip Leslie, Richard Mortimer, Ariel Pakes, and Frank Wolak for helpful discussions and comments. We also thank seminar participants at the Stanford Institute for Theoretical Economics, the National Bureau of Economic Research (Summer Institute), the North American Winter Meeting of the Econometric Society, Brown University, Columbia University, Harvard University, London School of Economics, the Stern School of Business at New York University, Northeastern University, Stanford University, University of Arizona, University of Texas, and Washington University for helpful comments. Financial support for this research was generously provided through the National Science Foundation (grant SES-0617896).

† Go to http://dx.doi.org/10.1257/mic.5.4.1 to visit the article page for additional materials and author disclosure statement(s) or to comment.

[1] Relatedly, firms throughout the economy have made large investments in technologies for tracking inventory and capacity information. For example, several large retailers require suppliers to use inventory tracking technology when distributing their products. Many other firms have adopted wireless communication and networked data centers for the same purpose.

forced substitution overstates demand for remaining goods conditional on the full-choice set being available. As a result, failing to account for product availability leads to biased estimates of demand substitution patterns, typically making products look more substitutable than they really are. This bias can potentially undermine the reliability of many important applications of demand estimates for markets with incomplete product availability, such as simulating the welfare implications of mergers or new product introductions, or applying antitrust policy. Identifying unbiased demand estimates in these markets is also a critical step in evaluating the optimal capacity choices or pricing strategies of firms.

In this paper, we provide evidence that failing to appropriately account for periods of product unavailability can result in a substantial bias in demand estimates, and we develop a method for correcting this bias. To accomplish this, we collect a new and extensive dataset with detailed inventory and sales information. The dataset covers one of the first technological investments for wirelessly managing inventory: a wireless network installed on a set of 54 vending machines, providing updates on elapsed sales and inventory status every four hours. The data from the vending network provide extremely granular information on sales and inventory levels over a one-year period. Using this dataset, we develop and implement estimation methods to provide corrected estimates even when some choice sets are latent, and we analyze the impact of stock-out events for firm profitability in the short run. We find evidence of important biases in predicted sales and substitution patterns for models that do not account for stock-out events in this market. In terms of the short-run impacts of stock-out events on profitability, our corrected model estimates that the negative profit impacts of stock-out events are up to 32 percent larger than the uncorrected model predicts.

The wireless vending network provides data on actual stock-out events, which randomly change the set of products available at some locations for a period of time. This variation provides an attractive source of identification for estimating demand models, because although the probability of a stock-out event can be targeted by the firm, the occurrence of any particular event is determined by the random order in which consumers arrive. Thus, stock-out events generate exogenous short-run variation in choice sets, in addition to the long-run variation that is more typically the source of identification for structural models of demand. Many other data sources have the potential to generate similar variation in availability, which makes the exercise here potentially quite broadly applicable. For example, markets for which capacity constraints are important, such as airplane flights or concert events, generate similar variation in choice sets when demand exceeds capacity.

When discussing inventory systems we use the standard language established by Hadley and Whitman (1963). The first of two types of inventory systems is called a "perpetual" data system. In this system, product availability is known and recorded when each purchase is made. Thus for every purchase, the retailer knows exactly how many units of each product are available.[2] The other type of inventory system is

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

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-out events because the consumer's choice set is not recorded with each transaction. While perpetual inventory systems are becoming more common in retailing environments, most retailers still do not have access to such systems. Sampling inventory more frequently helps to mitigate limitations of the periodic inventory system. However, an additional goal of this paper is to provide consistent estimates of demand not only for perpetual inventory systems but for periodic ones as well.

Indeed, despite the extremely detailed information in our dataset, we observe stock-out events only periodically (every four hours). Some stock-out events occur in the middle of an observed four-hour time period, and for these observations, the choice set of an individual consumer is latent. These latent stock-out events happen when initial inventory levels are low, but also when demand is high. Indeed, average sales during periods in which a product stocks out are nearly three times higher than sales during periods without stock-out events (15 versus 6.3 units). Thus, discarding data from these periods results in a sample that is selected on sales levels, and leads to a biased estimate of demand. While sales of remaining products are not observed before and after a latent stock-out event occurs for these periods, we show that the distribution of sales before and after such latent stock-out events is implied by the demand model. We develop a method for incorporating such periods into demand estimates that uses the well-known EM algorithm from the statistics literature (Dempster, Laird, and Rubin 1977) to estimate the allocation of sales across the unobserved choice-set regimes.

As technologies like the one we study continue to become more prevalent, firms and researchers can expect to gain access to more detailed information on sales and inventory/capacities, providing valuable information on short-run choice-set variation. Our results indicate that accounting for that choice-set variation can substantially reduce potential biases in standard demand estimates for some markets, and we conclude that researchers should take on the responsibility to adjust for the effects of product availability in demand estimation when possible.

Although not estimated here, the model we develop is also necessary for any examination of supply side decisions over the long run. For example, estimation of optimal capacity choices, restocking decisions, or inventory policies in markets where stock-out events matter, relies on a static demand model that accounts correctly for stock-out events as our model does. Demand estimates that correctly account for product availability are also important for understanding the macroeconomic implications of inventories. Indeed, firms' abilities to manage inventories have been proposed as an agent for dampening recessions, a factor affecting vertical relationships, and a strategic variable affecting price competition.

Relationship to Literature

The demand estimation literature in empirical industrial organization (IO) has primarily focused on cases in which all products are assumed to be available to all consumers (Berry, Levinsohn, and Pakes 1995; Nevo 2001; Berry, Levinsohn, and Pakes 2004). Dynamic models of demand (Hendel and Nevo 2006) have modeled consumer inventories or stockpiling, but do not incorporate data on availability at the time of purchase. Data on capacities/inventories have been used primarily in this literature in order to analyze supply side behavior—particularly with respect to promotions (Aguirregabiria 1999).[3]

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

Stock-out events 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 a six-product soft-drink machine and observe the inventory at the beginning of each day. The authors assume that each product is sold at a constant Poisson distributed rate that is independent from all other products. When a product stocks out, a new set of parameters is estimated with the restriction that the new predicted sales rates of remaining products are at least as large as the original rates.[4] Like us, the authors also address the issue of periodic inventory systems through the use of the EM algorithm to estimate which sales occurred before and after latent stock-out events. However, because of the lack of a utility-based framework for demand, the ADG method cannot be used to make out-of-sample predictions about alternative policies or their welfare impacts. More recently, Musalem et al. (2008) study the problem of stock-out events by imputing the entire sequence of sales in a Markov-Chain Monte Carlo approach. The MLE approach that we use is more efficient and faster to implement, particularly for markets in which the outside good has a relatively large share and the total number of products that stock out during periods of latent availability is relatively small.

Fox (2007) examines semiparametric estimation of multinomial discrete choice models using a subset of choices, and proposes a maximum score estimator. His focus includes settings with very large choice sets, for which estimation is computationally burdensome, as well as choice subsets that arise due to data limitations or consideration sets. However, this methodology does not allow for computation of market shares, which limits the ability to analyze the welfare effects of stock-out events.

The paper proceeds as follows. Section I provides the model of demand that adjusts for changes in product availability in the data under both perpetual and

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

[4] 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.

periodic inventory systems. In Section II, we provide estimation details and discuss identification of the model. Section III describes the data from the wireless vending route and provides correlations and regression results from the data. Section IV reports results from estimating the model using the vending data, Section V provides counterfactual experiments on the effect of stockouts on firm profitability, and Section VI concludes.

I. Model

Consider a discrete-choice model where consumer i chooses product j in market t. The choice probability is assumed to be a function of the set of available products, a_{it}, observable characteristics of the choice scenario, x_{jt}, and parameters of the choice model, θ.[5] This choice probability may result from a random utility maximization (RUM) problem such as in McFadden (1974), or Berry (1994), or from a simpler form such as a Poisson model. For each market t, sales are distributed multinomially with parameters M_t and p_{ijt}, and the (semi-nonparametric) log-likelihood contribution of a single market (where $y_{ijt} = 1$ if consumer i buys product j and 0 otherwise) is

$$l_t(\theta) = \sum_{i=1}^{M_t} \sum_{j=0}^{J} y_{ijt} \log p_{ijt}(a_{it}, x_{jt}, \theta). \tag{1}$$

A common parametric restriction is that $p_{ijt} = p_j(a_t, x_{jt}, \theta)$, (i.e., the choice probabilities are symmetric across consumers and markets conditional on x_{jt} and the choice set a_t). This restriction enables us to aggregate data across individuals within an (a_t, x_{jt}) duple, but does not rule out unobserved heterogeneity such as correlated tastes or random coefficients. For the multinomial distribution, total sales of each product under each (a_t, x_{jt}) duple are sufficient statistics for the likelihood, and we denote total sales from each market $y_{jt} = \sum_{i=1}^{M_t} y_{ijt}$, which can be represented as a vector $\mathbf{y}_t = [y_{0t}, y_{1t}, \ldots, y_{Jt}]$. Finally, when using aggregate data, the likelihood accounts for the fact that we do not observe the order of consumer arrivals through the multinomial coefficient:

$$C(\mathbf{y}_t) = \left(\frac{\left(\sum_{j=0}^{J} y_{jt}\right)!}{y_{0t}!\, y_{1t}!\, y_{2t}! \ldots y_{Jt}!} \right). \tag{2}$$

Thus, the log-likelihood written as a function of aggregate data is[6]

$$l_t(\mathbf{y}_t \,|\, \theta, a_{it}, x_{jt}) = \log\left(C(\mathbf{y}_t)\right) + \sum_j y_{jt} \log p_j(a_{it}, x_{jt}, \theta). \tag{3}$$

[5] x_{jt} may include product characteristics such as price, as well as characteristics of markets such as time-of-day.

[6] Note that the multinomial coefficient $\log\left(C(\mathbf{y}_t)\right)$ depends only on data and enters as an additive constant in the log-likelihood; thus it does not affect the choice of θ, and may be ignored in estimation.

A. *Incorporating Stock-Out Events*

Consider extending the model in (3) to allow a single good k to stock out, and define a_t and $a_{t'} = a_t \backslash k$ as the sets of products available to consumers before and after the stock-out event respectively. From (3), we define the sufficient statistics as the sales before and after the stockout: $y_{jt}^{a_t}$ and $y_{jt}^{a_t'}$, and construct a modified log-likelihood equation:

$$(4)\quad l_t(\mathbf{y}_t \,|\, a_t, a_{t'}, x_{jt}; \theta) \propto \sum_j y_{jt}^{a_t} \cdot \log p_j(a_t, x_{jt}; \theta) + y_{jt}^{a_t'} \cdot \log p_j(a_{t'}, x_{jt}; \theta).$$

In order to characterize how variation in the set of available products may be used to estimate θ, it is helpful to consider three types of markets. In the first type, availability is the same for all consumers, and we observe a_t and y_{jt}, so that $y_{jt} \equiv y_{jt}^{a_t}$ and $y_{jt}^{a_t'} \equiv 0$. This is the standard assumption maintained in much of the differentiated products literature. In the second type of market, availability varies across consumers, but is observed for each sale. Thus, $y_{jt}^{a_t}$ and $y_{jt}^{a_t'}$ are observed by the consumer and the researcher. In the third type of market, availability varies across consumers but is not observed by the researcher. In this case, $y_{jt}^{a_t}$ and $y_{jt}^{a_t'}$ are unobserved, although we observe their sum, denoted $y_{jt} \equiv y_{jt}^{a_t}$ and $y_{jt}^{a_t'}$. Perpetual collection of inventory data generates only the first two types of markets. Periodic collection of inventory data generates all three types of markets. The latent stock-out event in the third type of market presents a challenge for estimation because the sufficient statistics $y_{jt}^{a_t}$ and $y_{jt}^{a_t'}$ are not observed.

However, the multinomial distribution implies that sales before the stock-out event, $y_{jt}^{a_t}$, are binomially distributed:

$$(5)\quad y_{jt}^{a_t} \sim \text{Bin}(y_{jt}, \rho_{jt}^{a_t}),$$

where $\rho_{jt}^{a_t}$ is the probability that $y_{jt}^{a_t}$ sales of product j occur before product k stocks out. One may then derive a distribution for $\rho_{jt}^{a_t}$ as a function of the sales of the stocked-out product, y_{kt}, and the demand parameters, θ.[7]

B. *Deriving the Distribution of Unobserved Sales*

Product k stocks out only when its sales, y_{kt}, equal its initial inventory. Consider a single stock-out event of product k with M_t consumers in a market.[8] If we know the number of consumers, r_t, that arrive before the stock-out event and purchase a

[7] Standard demand estimation techniques assume all products are always available (i.e., that $\rho_{jt}^{a_t}$ has a degenerate distribution with mass at one). Other approaches, such as Bruno and Vilcassim (2008), assume that $\rho_{jt}^{a_t}$ is equal to the fraction of time j is in stock across other periods of the dataset. Such approaches are valid for cases in which the missing data do not depend on $\mathbf{y}$ (see Tanner and Wong 1987). However, they need not give consistent estimates in the case of stock-outs, where the missing data include sales. In other words, consistent estimation of demand implies a specific distribution for the missing sales as a function of observed sales, and substituting arbitrary distributions instead need not give consistent estimates of θ. We provide additional detail on this point in Section A of the Mathematical Appendix.

[8] We provide the extension to the case of multiple stock-out events in Section B of the Mathematical Appendix.

product other than product k, then we know that $(M_t - y_{kt} - r_t)$ consumers arrive after the stock-out event, and we can express $\rho_{jt}^{a_t}(r_t, y_{kt}, M_t; \theta)$ as

$$(6)\quad \rho_{jt}^{a_t}(r_t, y_{kt}, M_t; \theta) = \frac{r_t \cdot p_j(a_t, x_{jt}; \theta)}{r_t \cdot p_j(a_t, x_{jt}; \theta) + (M_t - y_{kt} - r_t) \cdot p_j(a_{t'}, x_{jt}; \theta)}.$$

Periodic data collection does not allow us to observe r_t. However, the demand model implies that r_t follows a negative binomial distribution, which describes the number of failures r_t (sales of all other products) until y_{kt} successes (sales of the stocked-out product) are observed. The negative binomial p.m.f. is

$$f(r_t, y_{kt}, p_k(a_t, x_{kt}; \theta)) = \frac{(y_{kt} + r_t - 1)!}{r_t!(y_{kt} - 1)!} p_k(a_t, x_{kt}; \theta)^{y_{kt}}(1 - p_k(a_t, x_{kt}; \theta))^{r_t}.$$

The demand model also implies that k stocked-out before M_t consumers arrived (i.e., $r_t + y_{kt} \leq M_t$). Thus we rescale $f(\cdot)$ by its CDF and denote the conditional p.m.f. as $h(\cdot)$:

$$(7)\qquad h(r_t; y_{kt}, p_k(a_t, x_{kt}; \theta), M_t) = \frac{f(r_t; y_{kt}, p_k(a_t, x_{kt}; \theta))}{F(M_t; y_{kt}, p_k(a_t, x_{kt}; \theta))}.$$

Knowing the distribution $h(r_t; y_{kt}, p_k(a_t, x_{kt}; \theta), M_t)$ provides enough information to compute the expectation of the sufficient statistic $E[y_{jt}^{a_t}]$ by integrating out r_t:[9]

$$(8)\quad E[y_{jt}^{a_t} | y_{jt}; \theta]$$

$$= y_{jt} \sum_{r_t=0}^{M_t - y_{kt}} \frac{r_t \cdot p_j(a_t, x_{jt}; \theta)}{r_t \cdot p_j(a_t, x_{jt}, \theta) + (M_t - y_{kt} - r_t) \cdot p_j(a_{t'}, x_{jt}; \theta)} h(r_t; y_{kt}, p_k(a_t, x_{kt}; \theta), M_t)$$

$$E[y_{jt}^{a_t'} | y_{jt}, \theta] = y_{jt} - E[y_{jt}^{a_t}].$$

There are two key advantages to this approach. The first is that one need not consider sales of products, other than product j and the stocked-out product k, when computing the expected sufficient statistic. The second is that the expectation is a single summation, so there is no need to keep track of every possible value of $y_{jt}^{a_t}$.

[9] The unconditional probability of facing a stock out is given by $E_h\left[\frac{r_t}{M_t - y_{kt}}\right]$.

Knowing $E[y_{jt}^{a_t} | y_{jt}, \theta]$ allows us to compute the expectation of the log-likelihood over the missing data, which is linear in the sufficient statistics:

$$(9) \qquad E[l_t(y_t | a_t, x_{jt}; \theta)] = E\Big[\sum_j y_{jt}^{a_t} \ln p_j(a_t, x_{jt}; \theta) \,|\, y_t, \theta\Big] + E\Big[\sum_j y_{jt}^{a_t'} \ln p_j(a_{t'}, x_{jt}; \theta) \,|\, y_t, \theta\Big]$$

$$= E[y_{jt}^{a_t} | y_t, \theta] \ln p_j(a_t, x_{jt}, \theta) + E[y_{jt}^{a_t'} | y_t, \theta] \ln p_j(a_{t'}, x_{jt}, \theta).$$

Direct maximization of this likelihood is quite difficult for two reasons. First, one must compute an integral for every j when a stock-out event occurs, at each evaluation. Second, the "data" $y_{jt}^{a_t}$ depend on the parameters θ through the nonlinear functions $p_j(a_t, x_{jt}; \theta)$ and $p_k(a_t, x_{kt}; \theta)$, which introduces nonconvexities and complicates the gradient of the log-likelihood.

Fortunately, one can avoid direct maximization and use the Expectation-Maximization (EM) algorithm for estimation instead. The convergence of the EM Algorithm was established broadly by Dempster, Laird, and Rubin (1977), but much earlier for the multinomial distribution (Hartley 1958).

C. *An EM Algorithm*

For members of the exponential family, the EM algorithm is an iterative procedure that alternates between computing the expected log-likelihood for a given value of the model's parameters (the E-step) and maximizing the expected log-likelihood (the M-step). For the stock-out problem, the E-step is established by substituting the expected sufficient statistics, $E[y_{jt}^{a_t} | y_{jt}, \theta]$ and $E[y_{jt}^{a_t'} | y_{jt}, \theta]$, into the log-likelihood. Thus, we compute

$$(10) \qquad E[l(y_t | a_t, x_t; \theta^l z)] = E\big[y_{jt} | y_t, \theta^l\big] \ln p_j(a_t, x_{jt}; \theta^l).$$

For the M-step, one maximizes the expected log-likelihood function

$$(11) \qquad \hat{\theta}^{l+1} = \arg\max_{\theta} \sum_{j,t} E\big[y_{jt} | y_t, \hat{\theta}^l\big] \ln p_j(a_t, x_{jt}; \theta).$$

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

There is no additional integration in the M-Step, so optimization routines treat the imputed values as if they are data. Thus, researchers can use "off-the-shelf" routines for ML (or MSL, depending on the specification of $p_j(\cdot)$). The expectation over the missing data does not enter the optimization procedure. This reduces the complexity of the problem because the expectation is evaluated only once per major iteration rather that at each likelihood evaluation.

In general, there are some criticisms of EM-type procedures. The first is that EM algorithms are slow to converge (that is, they require many major iterations). This is true, and the rate of convergence depends on the amount of data that are missing and the degree to which the likelihood is affected by the missing data. For many cases involving stock-out events, both of these tend to be small. That is, the bulk of the data is nonmissing (stock-out events are infrequent), and for cases where we do observe them, choice probabilities do not change drastically except for close substitutes. The other problem that EM-type procedures generally have is that they are local optimizers (as are almost all nonlinear search routines used by economists) and require good starting values. Indeed, EM algorithms are slightly more dependent on starting values than other algorithms, but this is a problem that is endemic to the entire class of extremum estimators.

II. Estimation

Our approach for handling stock-out events assumes only that the distribution of sales follows a multinomial process, and that we observe market-level data. It puts no other restrictions on the choice probabilities $p_{jt}(a_t, x_{jt}; \theta)$. In this section, we present several common parameterizations of $p_{jt}(a_t, x_{jt}; \theta)$. An approach common to the marketing literature is to parameterize $p_{jt}(\cdot)$ in an semi-nonparametric way:

$$p_{jt}(a_t, x_{jt}; \theta) = \lambda_{j,a_t}.$$

The ML estimate of λ_{j,a_t} is the mean market share of product j under availability set a_t. With enough data, one might also condition on x_{jt}.[10] The advantages of this specification are that one avoids placing strong parametric restrictions on substitution patterns, and the M-Step is easy. The disadvantages are that one must estimate J additional parameters for each observed choice set, and forecasting is infeasible for choice sets that are never or rarely observed in the data. Furthermore, the lack of a utility-based framework means that out-of-sample predictions about alternative policies cannot be made.

The typical approach in the empirical IO literature is to specify a random utility maximization (RUM) problem. Assume that consumer i receives utility $u_{ijt}(\theta)$ from product j in market t, and chooses a product to solve

$$d_{ijt} = \arg\max_{j \in a_t} u_{ijt}(\theta) \tag{12}$$

$$u_{ijt}(\theta) = \delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2) + \varepsilon_{ijt},$$

where d_{ijt} indicates that consumer i chose product j in market t, $\delta_{jt}(\theta_1)$ denotes the mean utility of product j in market t, and $\mu_{ijt}(\theta_2)$ and ε_{ijt} are random components. We consider two common assumptions for the random components: a random coefficients logit and a nested logit. The random coefficients logit assumes that μ_{ijt}

[10] Anupindi, Dada, and Gupta (1998) take this approach in their treatment of stock outs.

includes normally distributed tastes for product characteristics and ε_{ijt} follows a Type I extreme value distribution. The nested logit assumes that $(\mu_{ijt} + \varepsilon_{ijt})$ follows a Generalized Extreme Value (GEV) distribution.

For the random coefficients model, we specify choice probabilities as

$$(13) \qquad p_{jt}(\delta_j, \sigma, \xi; a_t) = \int \frac{\exp\left[\delta_j + \sum_l \sigma_l \, \nu_{il} \, z_{jl} + \xi_t\right]}{1 + \sum_{j \in a_t} \exp\left[\delta_j + \sum_l \sigma_l \, \nu_{il} \, z_{jl} + \xi_t\right]} f(\nu_i),$$

where δ_j is a product fixed effect, and ν_{il} represents random tastes for each characteristic z_l of product j. We also include a market-level demand shifter ξ_t, which captures variation in demand for the outside good. To compute the integral in (13), a number of approaches can be employed; we use the sparse grids quadrature rules of Heiss and Winschel (2008).

Similarly, for the nested logit model we specify choice probabilities as

$$(14) \qquad p_{jt}(\delta_j, \lambda, \xi; a_t) = \frac{e^{(\delta_j + \xi_t)/\lambda_l} \left(\sum_{k \in g_l \subset a_t} e^{(\delta_k + \xi_t)/\lambda_l} \right)^{\lambda_l - 1}}{1 + \sum_{\forall l} \left(\sum_{k \in g_l \subset a_t} e^{(\delta_k + \xi_t)/\lambda_l} \right)^{\lambda_l}}.$$

This again includes product- and market-level fixed effects δ_j and ξ_t, as well as category-specific nesting parameters $\lambda_l \leq 1$. We also estimate a restricted model where all of the nesting parameters are equal, which is similar to the nested logit model of Berry (1994) or Cardell (1997). Both parametrizations allow one to easily predict choice probabilities as the set of available products changes by simply adjusting a_t.

A. *Identification with High Frequency Retail Data*

The objective of any discrete-choice model of demand is to understand substitution patterns, which requires variation in a_t and/or x_{jt}. The inclusion of both product- and market-level fixed effects in x_{jt} allows for quite flexible estimation of mean product quality. In addition, we exploit short-run changes in a_t that are induced by stock-out events, which generate quasi-experimental variation among available choice sets.[11] In our empirical application, prices are fixed within each product, which means that variation in a_t, rather than price variation, becomes the primary source of identification. This makes our application a clean setting for understanding the relationship between stockout-based choice-set variation and demand parameters. However, one cannot identify a price parameter in our application, even though our method

[11] Note that this identification strategy differs from approaches used for aggregate data that are observed over long periods of time (e.g., Berry, Levinsohn, and Pakes 1995 or Nevo 2001). In such cases, choice-set variation comes solely from long-run changes in the mix of products, such as the introduction of new brands, and changes in product characteristics or prices.

may be applied to markets that include price variation, for which one estimates a price coefficient.[12]

Another important aspect of high-frequency retail data is that the relevant market size, M, is often small. For example, in our vending application, we assume a market size of 20 consumers per machine-hour, roughly corresponding to the peak transaction throughput of a vending machine. GMM approaches assume that $M \to \infty$ in order to invert the market-share equation and accurately isolate the random components of utility without error. This is generally not a realistic assumption in small markets, which is why we use maximum-likelihood instead of GMM for our application. Sampling error in market shares that is due to measurement error in small markets is incorporated into the likelihood function. If instead one has large markets with stock-out events (e.g., Amazon.com), a GMM approach may be substituted for ML in the M-step.[13]

III. Industry Description, Data, and Reduced-Form Results

A. *The Vending Industry*

The vending industry is well suited to studying the effects of product availability in many respects. Product availability is well defined: goods are either in stock or not (there are no extra candy bars in the back, on the wrong shelf, or in some other customer's hands). Likewise, products are on a mostly equal footing (no special displays, promotions, etc.). The product mix and layout of machines is relatively uniform across the machines in our sample, and for the most part remains constant over time. Thus most of the variation in the choice set comes from stock-out events, which are a result of stochastic consumer demand rather than the possibly endogenous firm decisions to set prices and introduce new brands.[14]

Typically, a location seeking vending service requests sealed bids from several vending operators 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 percent of gross sales. Delivery, installation, and refilling of the machines are the responsibility of the vending operator, who chooses the interval at which to service and restock the machine, and collects cash. The vending operator is also responsible for any repairs or damage to the machines. The vending client will often specify the number and location of machines.

[12] Our estimation method allows for inclusion of a price instrument by either the addition of distributional assumptions to the ML problem, or the use of a GMM procedure. Draganska and Jain (2004) develop a method for including IVs and supply side restrictions into an ML estimator, assuming a normal distribution on the unobservable product attributes. For a GMM procedure, Berry (1994) and Berry, Levinsohn, and Pakes (1995) can be used in the M step, because these estimators improve the likelihood at each step. However, these methods rely on an assumption that individual markets are large ($M_t \to \infty$), which might be problematic in very granular datasets. For the most "extreme case" of granularity, Chintagunta, Dubé, and Goh (2005) provide a method to extend the Berry (1994) method with price IVs to cases with individual-level data. Finally, Yang, Chen, and Allenby (2003) provide a Bayesian estimator when IVs are required.

[13] Dempster, Laird, and Rubin (1977) provides the extension of the EM algorithm for GMM.

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

Vending operators may own several "routes," each administered by a driver. Drivers are typically paid partly on commission to encourage them to 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 features of the industry have motivated advances in data collection, which enable operators to not only monitor their employees, but also to transparently provide commissions to their clients and make better restocking decisions.

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

B. *Data Description*

In order to measure the effects of stock-out events, we use data from 54 vending machines on the campus of Arizona State University (ASU). This is a proprietary dataset acquired from North County Vending with the help of Audit Systems Corporation (see North County Vending 2004). The data were collected from the spring semester of 2003 and the spring semester of 2004. The ASU route was one of the first vending routes to be fully wireless enabled and monitored through Audit System's software. The wireless technology provides additional inventory observations between service visits by transmitting each machine's internal DEX data to the vending operator's warehouse several times per day (approximately every four hours).

The dataset covers snack and coffee machines; we focus on snack machines. Throughout the period covered by our dataset, the machines stock chips, crackers, candy bars, baked goods, gum/mints, and a few additional products. Some products are rarely present, and others are not food items, and we exclude such products from our analysis on the assumption that they are substantially different from more typical snack foods.[15] For a few brands of chips, we observe rotation over time in the same slot of the machine, and for these goods, we create two composite chip products (Misc. Chips 1 and Misc. Chips 2).[16] Finally, we combine two different versions of three products.[17] The 44 products in the final dataset are listed in Table A1 in the Appendix.

[15] Of these products, most are nonfood items, such as condoms, gum, and mints. The three food items we exclude have fewer than a dozen sales over all machines (i.e., Grandma's Lemon Cheese, Grandma's Chocolate Croissant, and Nestle 100 Grand).

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

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

Sales at each machine are observed at roughly four-hour intervals. Retail prices are constant over time, machines, and across broad groups of products. Baked goods typically vend for $1.00, chips for $0.90, cookies for $0.75, and candy bars for $0.65. As compared to typical studies of retail demand and inventories (which often utilize supermarket scanner data), there are no promotions or dynamic price changes. This means that in our application, price effects are not identified once product or category dummies are included. 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.[18]

The dataset also contains information on stock-out events and marginal cost data (the wholesale price paid by the vending operator) for each product. We report an upper and lower bound on the percentage of time in which a product is observed to have stocked out. The lower bound assumes that the product stocked-out at the very end of a 4-hour period of observation, and the upper bound assumes that it stocked-out at the very beginning of a 4-hour period. For most categories and products, this ranges from two to three percent, with larger rates of stock-out events for pastry items and pretzels. 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. Variable markups tend to be lowest on branded candy bars (about 50 percent), and high on chips (about 70 percent).

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

Table 1 summarizes the incidence of stock-out events across the four-hour machine-periods. There are a total of 44,458 such "markets" in the dataset. Of these, 39,552 do not experience any stock-out events, while 4,906 experience between one and three stock-out events. The periods containing stock-out events comprise 11 percent of all observations, but 22.7 percent of total sales. The average sales rate during periods with stock-out events is nearly three times higher than the sales rate during periods without stock-out events (15 versus 6.3 units).

C. *Reduce-Form Results*

Before applying the estimation procedure described above to the dataset, we first describe the results of two simple reduced-form analyses of stock-out events. In Table 2, we compute variable profits for each four-hour wireless time period at each individual machine (for a total of 44,562 observations) and regress this on the number of products stocked-out and machine fixed effects. The first specification

[18] For products with more than one serving per bag, the characteristics correspond to the entire contents of the bag.

Table 1—Breakdown of Markets by Incidence of Stock-Out Events

Stock-out events	Machine-period observations	Total sales
0	39,552	249,047
1	4,050	54,741
2	687	14,049
3	169	4,617
Total	44,458	322,454

Note: An observation is a four-hour period at a machine.

Table 2—Regression of Profit on Stock-Out Variables

	Specification		
	(1)	(2)	(3)
Products stocked-out	−1.000***	−0.609***	
	(0.0725)	(0.175)	
Maximum category stock outs		−0.876**	
		(0.357)	
Pastry stock outs			−1.243***
			(0.215)
Cookie stock outs			0.570
			(0.490)
Chips stock outs			−1.465***
			(0.212)
Chocolate stock outs			−1.317***
			(0.410)
Candy stock outs			0.145
			(0.479)
Constant	23.47***	23.72***	23.57***
	(0.207)	(0.230)	(0.215)
Observations	44,562	44,562	44,562
R^2	0.0800	0.0801	0.0804

Notes: An observation is a four-hour period at an individual machine (recorded wirelessly). Dependent variable is variable profit for a four-hour period. All specifications include machine fixed effects.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

(column 1) estimates the four-hour profit loss to be about $1.00 per product stocked-out with a base level of $23.47 in daily profits (at an average machine). Column 2 allows the effect of a stock-out event to differ based on the number of products stocked out in the category with the most missing products, in order to capture the fact that substitution to the outside good may increase when multiple products are unavailable in the same category (i.e., missing one candy bar and one brand of chips is different from missing two brands of chips). We estimate the effect of a stock-out event in the category with the most products missing to be about $1.48 per day, and the base effect of a product stocking out to be $0.61 per day. In column 3, we include the number of stocked-out products in each separate category. Stock-out events for pastries, chips, and chocolate candy result in a predicted profit loss of around \$1.24–\$1.47 per day, while the effects of stocking out of nonchocolate candy or cookies are not statistically different from zero.

All of these regressions suffer from endogeneity bias, and may be picking up many other factors, but they suggest some empirical trends that can be explained by the full model. First, stock-out events decrease hourly profits as consumers substitute to the outside good. Second, multiple stock-out events among similar products cause consumers to substitute to the outside good at an increasing rate, but in a way that is potentially heterogeneous across product categories.

Table 3 reports the results of a regression of stock-out rates on starting inventory levels. An observation is a product-service visit pair. We have 1,818 service visits in our dataset, and about 35 products in each for a total of 61,189 observations. We report results for Probit and OLS (Linear Probability Model) with and without product and machine fixed effects. We condition on the length of the service visit in all specifications. We find that an additional unit of inventory at the beginning of a service period reduces the chance of a stocking out by 0.8–1.6 percent. A full column of candy bars usually contains 20 units. This means that the OLS (fixed effects) probability of witnessing a stock-out event on a fully stocked candy product during one week for one machine is $25 - 1.59 \times 20 + 7 \times 1.77 = 5.5$ percent. For a product at a machine with a starting inventory of five units, the predicted chance of a stock-out event is about 30 percent. This provides some suggestive evidence that starting inventories can provide variation in stockout rates to help identify our demand model.

IV. Estimation Results

We estimate nested logit and random coefficients logit demand specifications using three different treatments for stock-out events. In the first treatment we assume full availability in all periods, including those periods in which a stock-out event was observed. Choice set variation in this specification is generated by the introduction or removal of products over time, and, to a lesser extent, from selective stocking of products in different machines. We refer to this as the "Full Availability" model, and it is the standard method of estimation in the literature. In the second treatment we account for stock-out events that are fully observed, but ignore data that are generated during periods in which the timing of a stock-out event was ambiguous. This generates a sample that is selected on sales levels, because we expect stock-out rates to be higher during periods of high demand. We call this the "Ignore" model. In the third treatment, we account for fully observed stock-out events, and use the EM algorithm to estimate which sales occurred under the various stock-out regimes within any ambiguous period.[19] This is the "EM-corrected" model.

[19] The data contain a small number of observations (less than one percent) in which more than three products stock out. The vendor believes that these events are likely due to data errors, or may indicate removal or replacement of a machine, and so we omit them from estimation. While including observations with many simultaneous stock-out events is possible for estimation, omitting them eases the computational burden in our application. For settings in which large numbers of products stock out within a period of observation, the reader may refer to the

Table 3—Regressions of Stock-out Rates on Starting Inventory

	OLS specifications			Profit specifications		
	(1)	(2)	(3)	(1)	(2)	(3)
Starting inventory	−0.790***	−1.232***	−1.593***	−0.882***	−0.870***	−0.990***
	(0.0173)	(0.0274)	(0.0297)	(0.0189)	(0.0223)	(0.0232)
Elapsed days	1.181***	1.175***	1.768***	1.062***	0.896***	1.274***
	(0.0411)	(0.0380)	(0.0477)	(0.0365)	(0.0295)	(0.0378)
Constant	16.45***	22.76***	25.01***			
	(0.337)	(0.439)	(0.464)			
Fixed effects	—	Product	Prod × Mach	—	Product	Prod + Mach
Observations	61,189	61,189	61,189	61,189	61,189	61,189
R^2/pseudo R^2	0.0445	0.1852	0.2893	0.0705	0.2410	0.2804

Notes: An observation is an individual product at a service visit for an individual machine. Dependent variable is the stock-out rate.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

We assume an overall rate of consumer arrivals of $M = 20$ per hour, roughly corresponding to the maximum throughput of a vending machine (processing about one transaction every three minutes).[20] In practice, actual sales levels at most machines will be substantially lower for most periods, but such heterogeneity in the overall marketsize is controlled for in the model by the market-level fixed effect ξ_t. The inclusion of ξ_t also implies that observations without any sales have no impact on the likelihood (i.e., they are perfectly predicted by ξ_t); therefore we exclude them from the sample. We estimate 44,458 such fixed effects after excluding periods with zero sales.

Table 4 reports the number of markets, the dimension of ξ_t, the number of observations (ie., product-market pairs), log-likelihood values, and parameter estimates from estimation of the nested logit and random coefficients specifications under each of the three treatments of stock-out events. The Ignore specification is estimated on only those markets in which availability is perfectly observed (39,552 out of 44,458 markets). The Full Availability and EM-corrected models are estimated on all markets.

Panels A and B report estimates from two nested logit specifications, a single nesting parameter, and five category-specific nesting parameters. Both nested logit models are estimated by full information maximum likelihood methods. We report the nesting parameter λ from McFadden (1978) rather than the σ correlation parameter from Cardell (1997) or Berry (1994).[21] Roughly speaking, $\lambda \approx (1 - \sigma)$ such that $\lambda = 1$ corresponds to a simple logit specification and $\lambda = 0$ corresponds to perfect correlation within group.

methods in Section C of the Mathematical Appendix on alternative computational methods, which avoid integration of the exact distribution.

[20] There are no instances for which sales of the outside good are zero or negative using this definition of M. There are some time periods which are very short, and we restrict the minimum number of consumers to correspond to a 90 minute period (i.e., $M =$ max (30 consumers, 20 consumers per hour)).

[21] We use FIML rather than the least-squares estimator for the nested logit because we worry both about the endogeneity of the within-group share, $\ln(s_{j|g})$, and about many small markets for which the only within-group sales are the sales of product j (i.e., the case of extreme measurement error).

Table 4—Nonlinear Parameter Estimates

	Full availability	Ignore	EM-corrected
Panel A. Single nesting parameter (λ)			
Category	0.583	0.697	0.704
	(0.027)	(0.017)	(0.030)
Negative log-likelihood	2,036,554	1,595,335	2,019,821
Panel B. Category-specific nesting parameter (λ)			
Pastry	0.540	0.665	0.650
	(0.090)	(0.118)	(0.102)
Cookie	0.624	0.767	0.753
	(0.053)	(0.091)	(0.079)
Chips	0.683	0.840	0.823
	(0.009)	(0.022)	(0.019)
Chocolate	0.630	0.770	0.758
	(0.012)	(0.028)	(0.024)
Candy	0.663	0.820	0.799
	(0.031)	(0.057)	(0.050)
Negative log-likelihood	2,035,479	1,594,366	2,018,494
Panel C. Random coefficients			
Calories	0.752	0.764	0.777
	(0.017)	(0.019)	(0.015)
Sugar	0.351	0.413	0.437
	(0.103)	(0.086)	(0.063)
Negative log-likelihood	2,036,395	1,595,248	2,019,475
Markets	44,458	39,552	51,752
Dimension of ξ_t	44,458	39,552	44,458
Observations	4,179,680	3,639,212	4,179,680

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

Comparing the Full Availability and Ignore models, the point estimates of the nonlinear parameters change significantly in most cases. The within-nest correlation in tastes for the two nested logit models is roughly $(1 - \lambda)$. Estimates from the Full Availability model ($\lambda = 0.58$) and the Ignore model (where $\lambda = 0.70$) imply within-nest correlation of 0.42 and 0.30 respectively. The EM-corrected estimate of λ, is not significantly different from the Ignore model. Standard errors are provided for all estimates.[22] In general, all parameters are estimated fairly precisely in spite of the large number of market fixed effects. The estimated parameter values are significantly different between the Full-Availability model and the two models that adjust

[22] Standard errors for the Full Availability and Ignore models are readily available. The EM-corrected model requires a correction to the usual standard errors to account for the fact that sales in periods of latent availability regimes are estimated. We provide this correction in Section D of the Mathematical Appendix.

for stock-out events. The EM-corrected model produces the most efficient results by incorporating more markets in estimation (in spite of the correction to account for estimating sales during periods of latent availability).

Panel B of Table 4 reports similar patterns in the nested logit specification with five nesting parameters. In this specification, the estimated correlation between products in the same nest is also higher under Full Availability (between 0.46 and 0.32) than under Ignore or EM-corrected (between 0.35 and 0.16). Panel C of Table 4 reports estimates from a random coefficients logit specification, in which random coefficients are estimated for each of two observable product characteristics: calories and sugar.[23] The random coefficients specification does not show significant differences in the nonlinear parameter estimates.

V. Estimated Sales and the Impact of Stock-Out Events

All of the nonlinear parameters are scaled by the normalizations for ξ_t and δ_j, so simply comparing their point estimates does not reveal how the models differ in terms of their predicted sales or substitution patterns. In this section, we use the results from the category-specific nested logit and the random coefficients logit specifications to predict sales and the impact of stock-out events on firm profitability. These predictions give an indication of how important the corrections to the demand system are likely to be for determining supply side decisions about capacity and restocking efforts (or more generally, in calculating the welfare impacts of mergers or the value of new products when capacities are important).

A. *Estimated Sales and the Distribution of* ξ_t

We begin by defining a "typical" machine as one containing the 35 most widely carried products (measured across machines and over time) out of the full set of 44 products in the data. For this typical machine, we predict sales for two different types of markets under the nested logit and random coefficients logit specifications. First we predict sales for a market with the median value of ξ_t in the subset of the data with no latent stock-out events (i.e., the 39,552 markets used to estimate the Ignore specification). We replicate the same exercise for the markets with latent stock-out events, using the median value of ξ_t from these markets. The Ignore model drops these markets in estimation, and so we discuss only the Full Availability and EM-corrected results for these markets.

For markets without latent stock-out events, adjusting for the observed set of available products under either the Ignore or EM-corrected models results in predicted total sales that are 9–13 percent higher than Full Availability predicts, with very similar outcomes under both the random coefficient and nested logit specifications. The censoring bias dominates: for all but one product, the Ignore and

[23] We observe two additional discrete product characteristics: indicators for cheese and chocolate. These are excluded from estimation because they were not identified after the inclusion of product dummies. We believe the nonidentification of these parameters in our particular setting is due to the lack of additional product characteristics that vary continuously, such as price. Such a characteristic is a key assumption more generally for identification (see Berry and Haile 2008).

EM-corrected models adjust sales upward. The nested logit specification provides evidence that the forced substitution bias dominates for one product (Rice Krispies), with a downward adjustment to predicted sales under the Ignore and EM-corrected models. Both the Ignore and EM-corrected models predict higher sales for those products that have the highest incidence of stock-out events.[24]

Predictions of sales for the median market, in which a stock-out event did occur, show that the selection effect (stronger demand leading to stock-out events) is quite strong. Predicted total sales are almost three times larger than predicted sales in the median market without stock-out events. The Ignore model drops these observations altogether, and so cannot predict sales in these markets. The EM-corrected model predicts sales that are about 10 percent higher than the Full Availability model. Again, the adjustment in the predicted sales of each product is strongly correlated with products' stock-out frequencies.[25]

To further understand the selection issue that arises from ignoring markets with latent stock-out events, we provide histograms of the distribution of the market fixed effects from all four model specifications in Figure 1. Markets with stock-out events are clearly a selected sample of markets. For example, the distribution of ξ_t under the EM-corrected nested logit specification has a mean of -0.12 for markets without stock-out events and a mean of 1.01 for markets with stock-out events. The gap is approximately equal to one standard deviation of ξ_t for all four model specifications.[26]

Finally, note that the almost 45,000 market fixed effects provide an unusual degree of flexibility in our application. They also lead to larger standard errors than one would estimate in more typical models without such a large number of fixed effects. We have estimated all specifications using two alternative fixed-effects specifications: machine-level fixed effects, and no fixed effects. The specification with machine-level fixed effects gives similar estimates for the nonlinear parameters and predicted sales as we report in our baseline results. The specification with no fixed effects produces estimates of the nonlinear parameters that are inconsistent with utility maximization for both the Full Availability and Ignore models.

B. *The Impact of Stock-Out Events*

In order to demonstrate how the different demand estimates affect the impact of stock-out events on profitability, we conduct an experiment in which we consider sales at the same typical machine and compare them to a machine where the two best-selling products in each category are unavailable.[27] For each of the removed products, we calculate the number of forgone sales predicted by the random

[24] Regressing the difference in predicted sales (EM-corrected–Full Availability) on the upper bound of stock-out frequencies for each product (and a constant term) produces a positive and significant coefficient on stock-out frequency of 0.07, and an R^2 of 0.87 and 0.84 for the random coefficient and nested logit specifications respectively.

[25] The full set of results at the product level are available in an online Appendix.

[26] The distribution of ξ_t for markets without stock-out events has large mass where market size corresponds to a four-hour period (i.e., $M = 80$) and total sales in the market are equal to one unit.

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

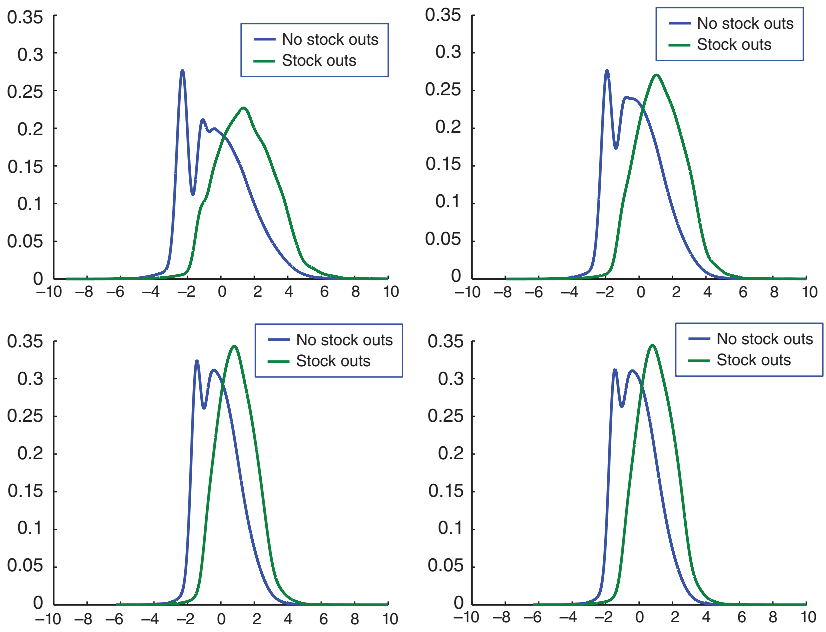

Figure 1. Distribution of ξ_t for Markets with and without Stock-Out Events

Notes: Reports the distribution of ξ_t for each model specification separately for markets without stock-out events (shifted to the left), and markets with stock-out events (shifted to the right). Model specifications are, clockwise from top left: Full Availability nested logit, EM-corrected nested logit, Full Availability random coefficient logit, and EM-corrected random coefficient logit.

coefficients and nested logit demand specifications. This ranges from roughly 8–12 in the chocolate category, 5–7 in the pastry and chips categories, and 3–5 in the cookie and candy categories.

We examine the results of the simulated product removals for the set of markets with no latent stock-out events. In both the random coefficients and nested logit specifications, the Ignore and EM-corrected estimates predict less substitution to available products than the Full Availability model, and lost sales that are 15–30 percent higher. This demonstrates the censoring and forced-substitution effects, because the removed products are the best-selling products in each category, and stock out more frequently in the data. The Ignore and EM-corrected models generally predict demand that is stronger for these best-selling, frequently stocked-out products (giving larger negative numbers of forgone sales for those products), and weaker for the remaining available products. We repeat the exercise for the approximately 4,900 markets in which a product stocked-out. Recall that we cannot make predictions for these markets using the Ignore model, so we compare the Full Availability model to the EM-corrected results. Lost sales are predicted to be 13 to 29 percent larger in the EM-corrected model under the two demand specifications.[28]

[28] The full set of results at the product level are available in an online Appendix.

Table 5—Weekly Profit Impact of Simulated Stock-Out Event

	Random coefficients			Nested logit		
	Full	Ignore	EM	Full	Ignore	EM
Panel A. Markets without stock-out events						
Foregone sales	−54.17	−62.21	−62.41	−55.97	−62.38	−62.62
Increased sales of substitutes	1.45	1.88	1.91	18.53	13.04	13.89
Overall change in sales	−52.72	−60.33	−60.51	−37.44	−49.34	−48.73
Percent staying inside	2.68	3.02	3.05	33.11	20.90	22.18
Change in profits	−26.93	−30.97	−31.10	−19.18	−25.31	−25.02
Panel B. Markets with stock-out events						
Foregone sales	−161.36	N/A	−183.82	−161.12	N/A	−179.97
Increased sales of substitutes	12.54	N/A	15.90	59.41	N/A	47.91
Overall change in sales	−148.82	N/A	−167.92	−101.72	N/A	−132.06
Percent staying inside	7.77	N/A	8.65	36.87	N/A	26.62
Change in profits	−75.79	N/A	−86.05	−51.91	N/A	−67.57

Notes: Reports predicted weekly profit impact of a simulated stock-out event of the top two selling products in each category (Chocolate Donuts, Strawberry Frosted PopTarts, Grandma's Oatmeal Raisin Cookie, Chips Ahoy Cookies, Rold Gold Pretzels, Sunchips Harvest Cheddar, Snickers, Twix, Starburst, and Kar Nut's Sweet and Salty Mix). Uses the same group of 35 most commonly held products from the sales and sales impact tables (see the online Appendix for details). The nested logit specification uses category-specific nesting parameters.

Finally, Table 5 summarizes the impact of the simulated stock-out events on profitability for the two types of markets. Panel A reports results for markets without stock-out events, while panel B addresses the roughly 4,900 markets in which a product stocked-out. The Full Availability model predicts lower levels of forgone sales compared to the Ignore and EM-corrected models under both demand specifications. The random coefficients specification predicts that more consumers choose the outside good, with roughly three percent staying inside versus 20–30 percent staying inside in the (category-specific) nested logit specification. The negative profit impact of the simulated stock-out events are predicted to be almost 11 percent larger for the random coefficients specification (roughly 32 percent larger for the nested logit specification) under the Ignore and EM-corrected models compared to Full Availability. A similar result holds when one compares the predicted profit impacts under the Full Availability and EM-corrected models for the markets in which a product stocked-out (i.e., a 13 percent increase for the random coefficient logit, and a 30 percent increase for the nested logit specification). For an industry with profit margins of less than four percent, these are economically significant differences.[29]

VI. Conclusion

Incomplete product availability is a common and important feature of many markets. This paper demonstrates that failing to account for product availability correctly can lead to biased estimates of demand, which can give misleading estimates of sales and the welfare impacts of stock-out events. We show that the

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

welfare impact of stock-out events in vending machines has a substantial effect on firm profits, indicating that product availability may be an important strategic and operational concern facing firms and driving investment decisions. Furthermore, biases that result from the incorrect treatment of stock-out events can potentially undermine the reliability of many important applications of demand estimates for markets with incomplete product availability, such as simulating the welfare implications of mergers or new product introductions, applying antitrust policy, constructing price indices, or evaluating the optimal capacity choices of firms.

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

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

Mathematical Appendix

A. *Treating the Stock-Out Time as a Free Parameter*

An alternative might be to consider the marginal data augmentation framework of Tanner and Wong (1987), in which we think of the fraction of consumers arriving before the stock-out event, α as the missing data and estimate it as an additional parameter. In general, this approach works when the integral is single dimensional because the integrand is a convex combination of choice probabilities. That is, there might exist a $\hat{\alpha}$ such that

$$\frac{\hat{\alpha} p_{jt}(\theta, a_t, x_{jt})}{\hat{\alpha} p_{jt}(\theta, a_t, x_{jt}) + (1 - \hat{\alpha}) p_{jt}(\theta, a_s, x_{jt})}$$

$$= \int \frac{\alpha p_{jt}(\theta, a_t, x_{jt})}{\alpha p_{jt}(\theta, a_t, x_{jt}) + (1 - \alpha) p_{jt}(\theta, a_s, x_{jt})} h(\alpha; \cdot) \partial \alpha.$$

If this were true, we could treat $\hat{\alpha}$ as an additional parameter to estimate. Unfortunately, we don't have a single equation, but rather a set of $(J - 1)$ equations (one for each product that did not stock out), and only a single α. Thus only in very special (degenerate) cases can a single $\hat{\alpha}$ satisfy all $(J - 1)$ equations.[30] This highlights the rationale for applying the E-step to the sufficient statistics, rather than some other quantity. In our case, the sufficient statistics are sales under each regime, rather than stock-out times. Our approach does follow the marginal data augmentation framework of Tanner and Wong (1987), but considers a model where we know sales under all availability sets (even though these aren't directly observed), rather than integrating the likelihood at each guess of the parameters.

Standard approaches do not solve (A1), but rather assume a different $f(\alpha)$ other than the one the stock-out distribution implies. For example, in the case where we ignore the missing data it would be as if we set $f(\alpha) = 0$ everywhere that things are ambiguous (which is not a proper density distribution). Or in the case of full availability it would be as if we set $f(\alpha)$ to be a delta function that took on value 1 only at the full availability value of a_t. An assumption that stock-out events happen at the beginning or the end of the period places similar structure on $f(\alpha)$ (making it a delta function). The problem with this is that $f(\alpha)$ does not have any free parameters, but is completely specified by the demand parameters as a conditional negative binomial.

B. *Multiple Unobserved Stock Outs*

Consider a market in which two products (labeled A and B) stock out. This creates four possible availability regimes a^0, a^A, a^B, a^{AB}, where superscripts denote the product or products that are stocked-out in any given regime. The sufficient statistics for product j are now distributed multinomially $\left[y_{jt}^0, y_{jt}^A, y_{jt}^B, y_{jt}^{AB}\right] \sim \text{Mult}(y_{jt}, \rho_{jt}^0, \rho_{jt}^A, \rho_{jt}^B, \rho_{jt}^{AB})$, where the lowercase subscript denotes the product whose sales we are considering. The form of ρ_{jt} is similar to the single stock-out case, and gives the probabilities of choosing j under the choice sets (a^0, a^A, a^B, a^{AB}), integrated out over the distribution of probabilities of facing each of the different choice sets. The challenge in this case is that the functional form of $h(\cdot)$ becomes more complicated.

To simplify things, consider the sales of the two stocked-out goods (Q_a, Q_b) and group all other sales into a composite good $Q_o = M_t - Q_a - Q_b$. The fraction of consumers of the composite good that face each choice set is given by

$$\alpha^0 = \frac{q_o^0}{Q_o}, \qquad \alpha^A = \frac{q_o^A}{Q_o}, \qquad \alpha^{AB} = (1 - \alpha^0 - \alpha^A).$$

If we assume that A stocks out before B then we can write the p.m.f. as a function of the sales of the composite good before either A or B stocks out $\left(q_o^0\right)$, and after

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

product A stocks out $\left(q_o^A\right)$, with an additional (nuisance) variable that determines how many sales of B took place before A stocked-out $\left(q_b^A\right)$. This is

$$h\left(q_o^0, q_o^A\right) = \sum_{q_b^0=0}^{Q_b-1} \underbrace{\Pr\left(Q_a, q_b^0, q_o^0\right)}_{\text{nmultpdf}\left(Q_a,\, q_b^0,\, q_o^0,\, p^0\right)} \cdot \underbrace{\Pr\left(Q_b - q_b^0, q_o^a\right)}_{\text{nbinpdf}\,\left(q_o^A,\, Q_b - q_b^0,\, p_b^A\right)}.$$

We evaluate the p.m.f. at every point in its support:

$$q_o^0 : \{0, \ldots, Q_o\}$$

$$q_o^A : \{0, \ldots, Q_o\} \quad s.t. \quad q_o^0 + q_o^A \leq Q_o.$$

The second term addresses the distribution of sales of the composite good after A has stocked-out. This is just the single stock-out case (for good B), which asks: What is the distribution of sales for the composite good when $\left(Q_b - q_b^0\right)$ sales of good B are realized? The first term addresses the initial stock out of product A, and asks: What is the joint distribution of the sales of $\left(q_o^0, q_b^0\right)$ when Q_a sales of A have occurred. This defines a negative multinomial distribution.[31]

The probability that A sells out before B is given by: $\Pr(q_a = Q_a, q_b < Q_b)$. We only care about consumers choosing A or B, so we can denote $\tilde{p} = \frac{p_b}{p_a + p_b}$ (the probability a consumer chooses B conditional on choosing A or B). This asks, "What is the probability of Q_A successes and less than Q_B failures?," and is the CDF of yet another negative binomial:

$$\begin{aligned}\Pr(A \succ B) &= \Pr(q_a = Q_a, q_b < Q_b \mid \tilde{p}) \\ &= \Pr(q_b < Q_b, N = Q_a + q_b \mid \tilde{p}) \\ &\sim \text{nbincdf}(Q_a - 1, Q_a, \tilde{p}).\end{aligned}$$

Finally, we condition on the event that both stock-out events occurred before M consumers arrived to the market, (i.e., $Q_o = M - Q_a - Q_b$), by normalizing $h(q_o^0, q_o^A) + h(q_o^0, q_o^B)$ to sum to one.

The extension to three or more stock-out events follows a recursive argument. Again we assume that A stocks out before B which stocks out before C so that the ordering is known. Again it is helpful to define $Q_o = M - Q_a - Q_b - Q_c$. Now we need the quantities $(q_o^0, q_o^A, q_o^{AB}, q_b^0, q_c^0, q_c^A)$.

[31] The p.m.f. of the negative multinomial is

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

The p.m.f. is

$$h(q_o^0, q_o^A, q_o^{AB}) = \sum_{q_b^0=0}^{Q_b-1} \sum_{q_c^0=0}^{Q_c-1} \sum_{q_c^A=0}^{Q_c-q_c^0-1} \underbrace{\Pr\left(Q_a, q_b^0, q_c^0, q_o^0\right)}_{\text{nmultpdf}\left(Q_a, q_b^0, q_c^0, q_o^0, p^0\right)} \cdot \underbrace{\Pr\left(Q_b - q_b^0, q_c^A, q_o^A\right)}_{\text{nmultpdf}\left(Q_b - q_b^0, q_c^A, q_o^A, p^A\right)} \cdot \underbrace{\Pr\left(Q_c - q_c^0 - q_c^A, q_o^{AB}\right)}_{\text{nbinpdf}\left(Q_c - q_c^0 - q_c^A, q_o^{AB}, p^{AB}\right)}.$$

We evaluate the p.m.f. at every point in its support:

$$q_o^0 : \{0, \ldots, Q_O\}$$

$$q_o^A : \{0, \ldots, Q_O\}$$

$$q_o^{AB} : \{0, \ldots, Q_O\} \quad s.t. \quad q_o^0 + q_o^A + q_o^{AB} \leq Q_o.$$

Finally, we ensure that the aggregate probabilities sum to 1 over all possible orderings of the stock-out events.

C. *Alternative Computational Methods*

For up to three unobservable stock-out events, the exact method we present in the text is generally computationally feasible. However, when many stock out at once, exhaustively evaluating over the distribution $h(\boldsymbol{\alpha} \mid \cdot)$, at all values of the domain becomes prohibitive, and approximate methods must be used. In this case, α represents the probability a consumer faces a particular choice set (for the single stock-out case $\alpha = \frac{r_t}{M_t - y_{kt}}$). Recall the form of the E-step:

$$E\left[y_{jt}^s\right] = \sum_{\forall s} y_{jt} \frac{\alpha_s \, p_j(\theta, a_s, x_{jt})}{\sum_{\forall r} \alpha_r \, p_j(\theta, a_r, x_{jt})} h(\boldsymbol{\alpha} \mid \theta, y_{jt}) \quad \text{s.t.} \quad \sum_{\forall r} \alpha_r = 1.$$

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

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

D. *Standard Errors: Missing Data Correction*

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

$$H(\theta) = -E\left[\frac{\partial^2 L(\theta\,|\,Z)}{\partial \theta^2}\right] = -E\left[\frac{\partial l(\theta\,|\,Z)}{\partial \theta}\,\frac{\partial l(\theta\,|\,Z)'}{\partial \theta}\right] = \left[\sum_i \frac{\partial l_i(\theta\,|\,Z)}{\partial \theta} \times \frac{\partial l_i(\theta\,|\,Z)'}{\partial \theta}\right]$$

$$l_i(\theta\,|\,y) = y_{jt}(\theta)\,\ln p_{jt}(\theta).$$

We obtain the score by differentiating the log-likelihood:

$$\frac{\partial l_i(\theta\,|\,y)}{\partial \theta} = \frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta}\,\ln p_{jt}(\theta) + \underbrace{\hat{y}_{jt}(\theta)\frac{1}{p_{jt}(\theta)}\,\frac{\partial p_{jt}(\theta)}{\partial \theta}}$$

$$= \sum_{jt} \frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta}\,\ln p_{jt}(\theta) + \frac{\partial l_i(\theta\,|\,\hat{y}_{jt})}{\partial \theta}.$$

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

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

$$\frac{\partial \hat{y}_{jt}(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\sum_{r_t \leq M_t} y_{jt} \frac{r_t \cdot p_{jt}(\theta)}{r_t \cdot p_{jt}(\theta) + (M_t - y_{kt} - r_t) \cdot p'_{jt}(\theta)} h(r_t | y_{kt}, p_k(\theta))\right]$$

$$= \frac{\widehat{y_{jt}^0}}{p_{jt}} \frac{\partial p_{jt}}{\partial \theta} + \widehat{y_{jt}^0} b_t \frac{\partial p_{lt}}{\partial \theta} - \left(\frac{\partial p_{jt}}{\partial \theta} - \frac{\partial p'_{jt}}{\partial \theta}\right) \frac{\widehat{y_{jt}^2}}{p_{jt}}$$

$$- (M_t - y_{jt}) \frac{\partial p'_{jt}}{\partial \theta} y_{jt} p_{jt} E_h\left[\frac{r_t}{D(r_t)^2}\right] + \frac{\partial p_{lt}}{\partial \theta} y_{jt} p_{jt} E_h\left[\frac{r_t^2}{D(r_t)}\right],$$

where

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

and

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

This expression indicates that: (i) the quadratic term dies out as gradient of the choice probabilities becomes more similar before and after the stock-out event (which we could think about as the finite difference of the stock-out event), and (ii) the other terms depend on how quickly the choice probabilities change in the parameters for the stocked-out product and the product of interest, roughly weighted by how much data is missing y_{jt}.[32]

Delta Method Correction for Nested Logit Parameters.—Finally, for the nested logit specifications, we require a correction for the standard errors for the product dummies, the d_js. We make the substitution $\tilde{\delta}_j = \frac{d_j}{\lambda_{j,k}}$ in estimation, where we define $\lambda_{j,k}$ as the nesting parameter which corresponds to the jth product. We write $\theta = [\tilde{\delta}_1, \ldots \tilde{\delta}_J, \lambda_1, \ldots, \lambda_K] \in \mathbb{R}^L$, and use the delta method to recover standard errors on the d_js. The asymptotic distribution for θ is

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

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

The asymptotic distribution of our normalized d_js is the asymptotic distribution of $g(\theta)$, which we define as

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

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

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

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

And the standard error is

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

For the λ parameters, no correction is necessary.

E. *Summary Statistics*

This section reports the full set of summary statistics by product, as discussed in Section IIIB.

Table 1A—Summary of Products and Markups

Product	Category	Stocked-out Lower/upper		Price	Cost	Share	Daily sales
PopTart	Pastry	6.32	7.49	1.00	0.35	3.73	1.41
Choc Donuts	Pastry	27.69	30.99	1.00	0.46	2.94	1.27
Ding Dong	Pastry	17.66	20.44	1.00	0.46	2.80	1.14
Banana Nut Muffin	Pastry	8.48	9.81	1.00	0.46	2.72	1.03
Rice Krispies	Pastry	1.72	2.09	1.00	0.31	2.01	0.78
Pastry	Pastry	20.40	23.09	1.00	0.46	0.84	1.58
Gma Oatmeal Raisin	Cookie/Snack	2.65	2.98	0.75	0.23	2.75	1.08
Chips Ahoy	Cookie/Snack	1.37	1.55	0.75	0.25	2.51	0.96
Nutter Butter Bites	Cookie/Snack	0.87	0.92	0.75	0.26	1.76	0.70
Knotts Raspberry Cookie	Cookie/Snack	0.97	1.05	0.75	0.19	1.76	0.68
Gma Choc Chip	Cookie/Snack	3.14	3.95	0.75	0.22	1.49	1.22
Gma Mini Cookie	Cookie/Snack	8.64	9.15	0.75	0.21	0.97	1.03
Gma Caramel Choc Chip	Cookie/Snack	11.30	12.14	0.75	0.22	0.41	1.05
Rold Gold	Chips	8.55	10.42	0.90	0.27	3.95	1.48
Sunchip Harvest	Chips	6.94	8.60	0.90	0.27	3.83	1.44
Dorito Nacho	Chips	2.79	3.54	0.90	0.27	3.36	1.26
Cheeto Crunchy	Chips	4.44	5.44	0.90	0.27	3.34	1.25
Gardetto Snackens	Chips	2.86	3.31	0.75	0.27	3.27	1.91
Ruffles Cheddar	Chips	5.78	7.19	0.90	0.27	2.74	1.04
Fritos	Chips	2.10	2.39	0.90	0.27	1.91	0.72
Lays Potato Chip	Chips	2.68	3.04	0.90	0.27	1.67	0.63
Misc. Chips 2	Chips	3.32	3.89	0.90	0.28	1.55	0.60
Munchies Hot	Chips	4.39	5.55	0.75	0.25	1.35	1.22
Munchies	Chips	7.13	8.18	0.90	0.25	1.18	0.93
Misc. Chips 1	Chips	5.98	6.75	0.90	0.28	1.18	0.82
Dorito Guacamole	Chips	3.80	4.19	0.90	0.28	0.91	0.73
Snickers	Chocolate	2.34	3.17	0.75	0.33	8.25	3.08
Twix	Chocolate	2.34	3.14	0.75	0.33	6.11	2.28
M&M Peanut	Chocolate	2.71	3.39	0.75	0.33	4.64	1.74
Reese's Cup	Chocolate	1.08	1.29	0.75	0.33	2.36	0.88
Kit Kat	Chocolate	0.84	1.11	0.75	0.33	2.16	0.81
Caramel Crunch	Chocolate	1.03	1.25	0.75	0.33	2.11	0.82
M&M	Chocolate	3.64	3.74	0.75	0.33	1.82	0.93
Hershey Almond	Chocolate	0.56	0.62	0.75	0.33	1.70	0.63
Babyruth	Chocolate	4.15	4.72	0.75	0.28	0.42	0.58
Starburst	Candy	5.00	5.76	0.75	0.33	3.18	1.43
Kar Nut Sweet/Salt	Candy	1.70	2.17	0.75	0.22	2.91	1.09
Snackwell	Candy	1.97	2.16	0.75	0.28	1.67	0.64
Skittles	Candy	1.73	2.23	0.75	0.34	1.53	1.21
Payday	Candy	0.56	0.62	0.75	0.33	1.18	0.83
Oreo	Candy	1.45	1.47	0.75	0.22	1.03	0.39
Peter Pan (cracker)	Candy	1.02	1.22	0.75	0.11	0.84	0.65
Peanuts	Candy	2.01	2.07	0.75	0.26	0.81	0.69
Hot Tamales	Candy	9.44	9.85	0.65	0.27	0.36	0.71

Notes: Product and category provided by the vending company, percent stocked-out (lower/upper bound) reports lower and upper bounds on stock-out frequencies over time, price is price charged at vending machines, cost is wholesale cost, share is 'inside good' market share, daily sales is average daily sales across all machines that carried a product.

REFERENCES

Aguirregabiria, Victor. 1999. "The Dynamics of Markups and Inventories in Retailing Firms." *Review of Economic Studies* 66 (2): 278–308.

Anupindi, Ravi, Maqbool Dada, and Sachin Gupta. 1998. "Estimation of Consumer Demand with Stock-Out Based Substitution: An Application to Vending Machine Products." *Marketing Science* 17 (4): 406–23.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25 (2): 242–62.

Berry, Steven T., and Philip A. Haile. 2008. "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers." http://www4.gsb.columbia.edu/filemgr?file_id=29144.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1): 68–105.

Bruno, Hernán A., and Naufel Vilcassim. 2008. "Research Note—Structural Demand Estimation with Varying Product Availability." *Marketing Science* 27 (6): 1126–31.

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13 (2): 185–213.

Chintagunta, Pradeep, Jean-Pierre Dubé, and Khim Yong Goh. 2005. "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models." *Management Science* 51 (5): 832–49.

Conlon, Christopher T., and Julie Holland Mortimer. 2013. "Demand Estimation under Incomplete Product Availability: Dataset." *American Economic Journal: Microeconomics*. http://dx.doi.org/10.1257/mic.5.4.1.

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 Series B* (*Methodological*) 39 (1): 1–38.

Draganska, Michaela, and Dipak Jain. 2004. "A Likelihood Approach to Estimating Market Equilibrium Models." *Management Science* 50 (5): 605–16.

Fox, Jeremy T. 2007. "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices." *RAND Journal of Economics* 38 (4): 1002–19.

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

Gupta, Sachin, Pradeep Chintagunta, Anil Kaul, and Dick R. Wittink. 1995. "Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data." *Journal of Marketing Research* 33 (4): 383–98.

Hadley, George F., and Thomson M. Whitman. 1963. *Analysis of Inventory Systems.* Prentice Hall International Series in Management and Quantitative Methods Series. Englewood Cliffs: Prentice Hall.

Hartley, H. O. 1958. "Maximum Likelihood Estimation from Incomplete Data." *Biometrics* 14 (2): 174–94.

Heiss, Florian, and Viktor Winschel. 2008. "Likelihood approximation by numerical integration on sparse grids." *Journal of Econometrics* 144 (1): 62–80.

Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica* 74 (6): 1637–73.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka, 105–42. New York: Academic Press.

McFadden, Daniel. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, 531–52. North-Holland.

Musalem, A., M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten. 2008. "Structural Estimation of the Effect of Out-of-Stocks." Columbia Business School Working Paper DRO-2008-12.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307–42.

North County Vending. 2004. "Sales data from Arizona State University route." www.vending.org (accessed using Audit Systems Corporation software).

Tanner, Martin A., and Wing Hung Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82 (398): 528–40.

Yang, Sha, Yuxin Chen, and Greg M. Allenby. 2003. "Bayesian Analysis of Simultaneous Demand and Supply." *Quantitative Marketing and Economics* 1 (3): 251–75.