Demand Estimation with Availability Variation¹

William Hickman (Boston College)

Julie Holland Mortimer (Boston College and NBER)

April 3, 2015

Estimates of demand are identified from variation in the choice sets that consumers face and the corresponding purchase probabilities for individual products. Retail settings often provide an opportunity to observe variation in consumer choice sets that arises not only through changes in observable product characteristics, such as price, but also through changes in product availability. We review the literature that develops methods for estimating demand in these settings, with emphasis on two mechanisms through which product availability may vary: product assortment decisions, and stockout events. We also briefly discuss variation in availability that may arise from limited consumer information.

1 Introduction

Demand estimation is fundamental to questions of economics and marketing because it informs optimal firm choices, and enables researchers to assess the social welfare effects of public policies or different market environments. Many common models of demand estimation – from individual household demand to aggregate market demand – assume that the set of available choices (henceforth the “availability set”) is the same for each agent. Notwithstanding its convenience in model construction, this assumption may be a poor approximation of reality for many retail environments. For example, Gruen, Corsten, and Bharadwaj (2002) report that the worldwide average retail stockout level of grocery and drugstore items is about 8% in developed countries. In other words, many consumers will encounter a stockout of a product they intended to purchase. In addition to stockouts, availability variation may also arise from differing assortments across stores in a market, or even differing awareness and consideration of products among consumers.²

Some variation in availability may plausibly arise exogenously and provide helpful sources of identification. However, ignored availability variation may also bias estimated demand and elasticities, resulting in misleading inferences about consumer and firm behavior, including optimal firm choices (such as prices, assortments, and inventories), and the effects of mergers and other policy changes.³ Model comparisons in both empirical applications and numerical simulations confirm that the biases are economically significant. Thus, we review several methods that account for variation in availability, examining their advantages and limitations.

The demand estimation methods we discuss differ along several dimensions, including the type of variation in availability, assumed data structures, underlying choice model, broad estimator type (e.g., maximum likelihood, hereafter ML) and academic field origin. Since we intend for this review to be

¹ We thank Chris Conlon, Sylvia Hristakeva, Tom Quan, and Kevin Williams for valuable feedback.

useful for applied researchers regardless of academic background, we have organized our study along two primary dimensions: type of variation in availability and assumed data structure; in particular, whether variation in availability is observed or unobserved. The three principle sources of variation in availability, ordered from most to least aggregated, are: assortment, stockouts (i.e., the outcome of capacity constraints and inventory levels), and limited consumer information (e.g., consideration sets).

Although there are many approaches to demand estimation, we restrict our review to discrete-choice demand.⁴ We assume the reader has basic familiarity with these models and their relationship to random utility maximization (RUM).⁵ Discrete-choice methods are now commonly used among economics and marketing scholars. They have been applied in settings with yearly, quarterly, or monthly observations covering either entire nations or large regions, as well as to observations covering weeks, days, or portions of a day in individual stores, chains of stores, or markets of limited geographic scope. At all these levels of aggregation across both time and space, product availability plays a role in the determination of observed sales.

Discussions of variation in availability, and biases in demand estimation more generally, easily venture into myriad other topics that are beyond the scope of this survey. For example, if consumers are willing to visit many stores in their market area, then variation in assortment is less problematic for demand estimation when aggregate market-level data are used, but more salient for store-level demand estimation. Despite the importance of store choice in some demand estimation settings, it lies beyond the scope of our survey here. The boundaries of our research also exclude such related subjects as consumer perception of variety, endogenous (observed) product characteristics, shelf space allocation, promotions, psychological effects of changes in availability (e.g. phantom effects), and dynamic demand. Finally, we exclude demand estimation procedures designed for a single product, rather than for a choice among multiple products.

In section 2, we review the prevalence of assortment variation and the assumptions needed to identify demand models, both when assortments are observed and when they are not. Section 3 discusses stockouts: the prevalence of stockouts, consequences of ignoring them, and models that account for unobserved stockout variation. We also briefly discuss the use of stockout-derived variables in demand models, experimental evidence of demand responses to stockouts, and related behavioral effects. We provide a short summary of models of limited consumer information in section 4. Section 5 concludes.

2 Assortment

An understanding of the role that product assortment plays in determining overall sales patterns is useful for both firms wishing to maximize profits and policy makers studying the effects of regulatory restrictions or other policy-relevant events. Interest in the effects of counterfactual assortments may be divided roughly into two groups along these lines. The first group, economists and policy makers, uses demand estimates to evaluate the welfare implications of, e.g., mergers or proposed policy changes. Demand estimates (usually along with a model of supply) provide a crucial input for predicting the effect

⁴ See Barnett and Serletis (2008) for a review of other approaches to demand estimation.

⁵ See Nevo (2011) or Train (2009) for introductions to discrete-choice demand estimation. Berry, Levinsohn, and Pakes (1995) develops the random-coefficients logit model, and Nevo (2000) provides further insight into its application.

of potential policy changes on the assortment of products provided in the market. The second group, business researchers and executives, is interested in the optimal firm choice as an end in itself.

Boundaries

Our focus in this section is the estimation of demand when the assortment of products varies, either across geographic or temporal markets, or across retailers within such a market. A number of scholars, particularly those in marketing and retailing, have examined related aspects of assortment changes, including the role of observable product characteristics of delisted brands in determining consumer decisions to switch stores or brands (Sloot and Verhoef (2008)), the effect of category assortment reductions on category sales (Boatwright and Nunes (2001)) or customers' shopping frequency and purchase quantity (Borle et al. (2005)), the difference between short- and long-term effects of assortment reductions (Sloot, Fok, and Verhoef (2006)), and the extent to which responses to stockout events explain responses to assortment reductions (Campo, Gijsbrechts, and Nisol (2004)).

We restrict our coverage of demand estimation to static models. Although we consider models in which assortment varies over time, we do not review models that are truly dynamic in the sense that consumers make purchase decisions that depend on anticipated future prices or assortments. Markets for durable goods, such as cars or computers, frequently require these dynamic models of demand. For example, Gowrisankaran and Rysman (2012) and Goettler and Gordon (2011) both provide models in which demand is affected by consumers' expectations of future assortments. Relatedly, Soysal and Krishnamurthi (2012) considers dynamic demand for seasonal goods.

Research on product lines and product variety can be found amongst both business and economics scholars (e.g., Berry and Waldfogel (2001), Broniarczyk, Hoyer, and McAlister (1998), Gourville and Soman (2005), Draganska and Jain (2005), Richards and Hamilton (2006), Misra (2008), Quan and Williams (2014), and Conlon and Mortimer (2015)). Many of these contributions are focused on modeling supply-side decisions, including the decisions of manufacturers to extend or shorten available product lines, and the decisions of retailers to stock particular product varieties. Related work also exists on consumers' choices of stores, though not directly as a function of assortment (e.g., Katz (2007) and Figurelli (2013)), as well as the interplay between market share and relative assortment share (Hwang and Thomadsen (Forthcoming)).

Updates to product lines (both by manufacturers and retailers) also affect economic policy. Pakes (2003) and Pakes and Erickson (2011) examine the importance of adjusting for changes in product assortment when calculating the Consumer Price Index (CPI). Pakes (2003) shows that using a hedonic approach to adjust for changes in observable product characteristics of entering and exiting products in the market for desktop computers has an important impact on the value of the CPI. Pakes and Erickson (2011) extends this method to the markets for televisions and digital cameras, for which the most important product characteristics are unobservable, and provide hedonic procedures that "at least partially account for the contribution of unobserved characteristics," using experimental variation.

The fields of Operations Research (OR) and Revenue Management (RM) have produced a theoretical literature on optimal assortment decisions by firms. Mantrala et al. (2009) provides a review of assortment planning from the retailing literature, Pentico (2008) surveys the work on optimal assortment in a variety of demand settings among OR scholars, and Kök, Fisher, and Vaidyanathan (2009) surveys

the OR and business (marketing and retailing) literatures on assortment, including a short discussion of demand estimation.

Before turning to the literature on structural demand estimation that accounts for heterogeneous assortment, we begin with an overview of the empirical evidence on assortment variation. These observations will motivate and inform the subsequent discussion of structural models and the assumptions required for their identification.

2.1 Prevalence of Assortment Variation

If assortment variation is sufficiently limited in practice, then researchers need not be concerned about ignoring it. Thus, it is useful to have a sense of the extent of assortment variation in common demand estimation settings. Using grocery store data, Tenn and Yun (2008) show that products with limited availability account for a significant portion of store revenues. They examine 132 weeks of grocery data from 14 retailer-city pairs for five product categories. Product availability, reported as a percentage known as the All Commodity Volume (ACV), is calculated as the fraction of all stores' annual sales, across all categories, accounted for by stores carrying the product. The authors divide products into three levels of availability: low (ACV between 0% and 50%), intermediate (between 50% and 90%); and wide (greater than 90%). Their results show that although many low-availability products constitute a small portion of sales (3% to 8% across the categories) and may be ignored without significant consequence, products of intermediate availability play a significant role in store sales (16% to 33%). A similar result emerges in work by Reibstein and Farris (1995), who find, after reviewing several papers, that market shares are convex in availability. Hwang and Thomadsen (Forthcoming) regresses national brands' market shares on, among other marketing mix factors, relative assortment share (i.e. number of stock-keeping units, or SKUs), finding a positive and significant relationship.

Empirical evidence of assortment variation can also be found in the descriptive statistics of several papers that have estimated demand in the context of assortment variation. In an application to chocolate confections in the United Kingdom, Bruno and Vilcassim (2008) report average marginal ACV-weighted availabilities for 24 products. They find a maximum (minimum) availability of 99% (14%), and a median availability of 80.5%. Quan and Williams (2014) show that the vast majority of shoes sold by Macy's in the U.S. are available at fewer than 10% of the chain's stores, and that the distribution of shoe availability across Payless stores is bimodal, with many products carried at fewer than 10% of stores, and many carried at nearly all stores. Misra (2008) reports variation in assortment across supermarkets within a city-chain pair for analgesics and coffee products, and finds that between 46% and 80% of products are not carried by all stores in a chain.

One might try to ameliorate the problem of assortment variation by aggregating to the product line, for which variation is substantially diminished. Tenn and Yun explain that such aggregation is unlikely to resolve the problem because of the importance of the underlying product composition to consumers. In order to treat product-line estimates as structural parameters one must assume that items within a product line are interchangeable as substitutes with other products and that product line lengths are identical across stores within each market. These assumptions are implausible in categories, such as grocery products, for which substitution patterns depend on flavors, colors, styles, sizes, and other characteristics. Furthermore, in some contexts, even product-line assortment varies significantly. Using data on skin-care products sold by retailers in a city of a developing country, where small- and medium- sized stores

account for about 90% of the ACV, Shah, Kumar, and Zhao (Forthcoming) demonstrates that the three largest brands each have average market availability of between 46% and 51%, despite collectively accounting for 89% of category sales. This suggests substantial assortment variation even at the product-line level.

While much of the variation in assortment exists across stores, variation within a store over time also occurs as new products are introduced, existing products are discontinued, and seasonal or promotion-related temporary changes to assortment begin and end. Tenn and Yun (2008) estimates that on average, about one-third of new products become widely available within a year, another third are completely discontinued, and the remainder continue to have an ACV less than 90%. They also find that, of the products that became widely available within a year, between 18% and 43% (depending on the category) become widely available within the first month, with an average time to wide availability of 8.6 to 14.2 weeks.

Products in rapidly growing categories and categories with frequent quality improvements feature rapid obsolescence of existing products and dramatic changes in assortment over time. In a study of the ready-to-drink tea product category, which grew by 700% over 5 years, Bronnenberg, Mahajan, and Vanhonacker (2000) reports brand ACV-weighted minimum (maximum) distribution rates ranging from 1% (9%) to 25% (71%) across brands. Eizenberg (2014) examines adoption rates in the CPU market (for which ‘adoption’ means that a product line offers the CPU as an option) and market shares for CPU product lines (e.g., Celeron, Pentium III, etc.). Across three years of quarterly data, within-product-line adoption rates commonly vary from single digit percentages to greater than 50%, reflecting the entry of new technologies and exit of old ones.

Hwang, Bronnenberg, and Thomadsen (2010) examines the drivers of assortment variation by regressing measures of assortment similarity between pairs of stores located in the same state on measures of similarity in four groups of variables: (i) store ownership, (ii) size, (iii) customer demographics, and (iv) competitive factors (e.g., number of nearby competitors and the distance between the stores). Their data contain one year of weekly observations on four supermarket categories (cola, ready-to-eat cereal, ground coffee, and toothpaste) from approximately 2,000 stores across 21 states. Their analysis suggests that store ownership variables have significant effects on assortment similarity, while measures of competitive factors are usually insignificant, both statistically and economically. The authors also compare the mean assortment similarities of firm-specific stores within a state and across states, finding that across-state similarities are significantly lower. Hwang et al. suggest this finding is driven by corporate organization of divisions that approximate state boundaries, and they conclude that their evidence is consistent with large costs of assortment variation relative to the gains of customization to local tastes.

Given the substantial level of sales from products of intermediate availability, and the consistent evidence of significant variation in assortment across time, stores, and geographic markets, researchers have cause to be concerned about bias in demand estimation that does not account for differing assortments. Hence, the next question to consider is ‘how large is this bias?’ Both unobserved and observed assortment variation can bias demand estimates and elasticities. We begin by discussing biases in demand models for unobserved assortment heterogeneity, and then review models designed to incorporate observed assortments.

2.2 Demand Estimation with Unobserved Assortment Variation

We review three papers that discuss the challenges of unobserved assortment variation for demand estimation from market-level data. Tenn (2009) examines sufficient assumptions for consistent estimates of own and cross-price elasticities in a linear demand model in product space. Tenn and Yun (2008) explore the sign and magnitude of assortment bias, both analytically using a heterogeneous-store logit model, and numerically through Monte Carlo simulations. Considered together, these papers form a set of negative results about the possibility of consistently estimating demand, especially cross-price elasticities, under certain conditions. They also provide empirical evidence that the conditions generating the negative results are likely to occur when using data aggregated above the individual-store level. Bruno and Vilcassim (2008) propose a method that accounts for assortment variation using market-level measures of distribution for each product, although it requires a strong independence assumption about the joint assortment distribution.

Bias in a Linear Demand Model

Starting with an underlying, but infeasible, model of store-specific parameters in a linear demand product-space model, Tenn (2009) presents a set of sufficient conditions that allow the model to be identified with aggregate data: (i) all prices in the market are the same, and (ii) the fraction of stores within a market that carry a particular assortment does not vary from one market to another.⁶ As Tenn points out, the single price assumption is unlikely to hold in many, if not most, cases, and the second assumption may be violated if assortment varies across different geographic markets.⁷

Combining the pricing and coverage assumptions with a third assumption — namely, that the quantity demanded of each product does not vary with the assortment offered in the store — one may write the probability limit of the estimated elasticities for the products contained in assortment g . If one uses aggregate data and a linear demand specification, this limit is

$$\bar{\varepsilon}_{jkm} = E_g(\varepsilon_{jkm}^g \mid j \in g, k \in g) \Pr(k \in g \mid j \in g) \quad (1)$$

where ε_{jkm}^g is the elasticity of demand for product j with respect to the price of product k for consumers facing assortment g , $\Pr$ represents probability, and subscript m is a market.⁸ The first term is the average elasticity of consumers who can choose either of the two products. Note that one cannot estimate this elasticity. Rather, estimation produces an attenuated elasticity due to the fact that in the data, aggregate sales of j come from stores with *and without* k in their assortment. This demonstrates that a special case of aggregate demand estimation produces cross-price elasticity estimates that are *not* the average elasticities for consumers who face a choice between the two products. One might be tempted to conclude from equation (1) that aggregate demand models always produce attenuated estimates of cross-price elasticities, but as Tenn and Yun (2008) point out, price coefficient bias induced by assortment variation

⁶ Alternatively one could write the demand model as a function of the average market price, and show that the unobserved price variation is part of the error term. This would introduce correlation between price and assortment, and possibly with store characteristics, and then one could consider whether or not it is possible to address those correlations.

⁷ Tenn notes that the second assumption may be more reasonable when a market is a distinct time period of a single geographic area and the temporal variation does not include significant seasonal variation in assortment.

⁸ Throughout this section, we harmonize notation such that markets are denoted as m , regardless of whether they refer to geographic or temporal markets.

can bias cross-price elasticities away from zero. One cannot conclude, in general, the direction in which assortment variation will bias elasticity estimates.⁹

Bias in a Multinomial Logit (MNL) Model

Tenn and Yun (2008) considers a standard (non-random coefficients) logit model of demand, for which the utility for consumer i of choosing product j in market m is given by $u_{ijm} = X_{jm}\beta - \alpha p_{jm} + \xi_{jm} + \epsilon_{ijm}$. In the utility function, X_{jm} is a vector of observed characteristics, p_{jm} is the price of j in market m , ξ_{jm} captures unobserved characteristics, and ϵ_{ijm} is the usual type 1 extreme value (TIEV) error that generates the logit form for the choice probabilities. This model results in the usual equations for own and cross-price elasticities:

$$\epsilon_{ijmr} = -\alpha p_{jm}(1 - s_{jmr}) \forall j \in J_{mr} \quad (2)$$

$$\epsilon_{jkmr} = \alpha p_{km} s_{kmr} \forall j, k \in J_{mr} : j \neq k \quad (3)$$

where ϵ_{jkmr} is the elasticity of demand for j with respect to price k in market m at retailer r , s_{jmr} is market share of j within retailer r of market m , and α is the price coefficient. As can be seen in equations (2) and (3), bias in the price coefficient will lead to bias in the estimated elasticities. Tenn and Yun's theoretical analysis reveals two potential sources of bias in the price coefficient — an omitted variable bias and bias from invalid price instruments — along with two additional sources of bias in estimates of elasticities — composition and weighting — all of which result from ignoring variation in assortment. We discuss each of these biases below.

In order to analytically demonstrate the omitted variable bias, Tenn and Yun assume that the expected market share of the outside good is independent of the availability of product j . This allows them to directly compare the price coefficient of a ‘homogeneous retailer’ model with that of a ‘heterogeneous retailer’ model. The homogeneous retailer model gives $\ln \frac{s_{jm}}{s_{om}} = X_{jm}\beta - \alpha p_{jm} + \xi_{jm}$, while the heterogeneous store model includes an additional additive term, $\ln \text{ACV}_{jm}$, where ACV_{jm} is the All Commodity Volume measure of product availability. Retail firms choose both assortment and price, and so correlation between price p_{jm} and $\ln \text{ACV}_{jm}$ is likely, biasing the estimate of α . In the experience of Tenn and Yun, price and product availability are most often negatively correlated, which implies that price coefficient estimates of the ‘homogeneous retailer’ model are biased upwards in absolute value.

In practice, many researchers use instruments for price when estimating the logit demand model. Supply-side cost instruments are often considered advantageous because, conditional on assortment remaining unchanged, they are unlikely to be correlated with unobserved components of demand.¹⁰ However, if

⁹ In their discussions of cross-price elasticity biases, both Tenn (2009) and Tenn and Yun (2008) focus on the elasticity for customers who can choose between the two products. Researchers may be interested in other elasticities, such as the elasticity for the entire market, including consumers for whom one of the two products is not available. Although the particularities of the bias will differ for those other elasticities, they will still depend on the unobserved joint distribution of assortments and fail to be invariant to assortment changes.

¹⁰ See, for example, Berry, Levinsohn, and Pakes (1995), among others.

supply-side costs affect both prices and the probability of being included in the assortment, then they are invalid instruments for price, because the error term contains ACV.

Elasticity estimates may also suffer from a composition bias. When assortment varies across retailers, the average elasticity of demand for k with respect to price j across the subset of consumers whose choice set includes both items is an expectation across markets and retailers given by

$$\varepsilon_{kj} = E_{mr}(\varepsilon_{kjm} | a_{kmr} = 1, a_{jmr} = 1)$$

where a_{kmr} is an indicator for availability of k in market m at retailer r . In contrast, for the homogeneous store model (i.e., without accounting for assortment variation), the analogous elasticity is given by

$$\tilde{\varepsilon}_{kj} = E_m(\tilde{\varepsilon}_{kjm} | j, k \in J_m)$$

where J_m is the set of products available in market m . Researchers often report only these average elasticities, because discrete-choice demand models feature many pairs of products, potentially in many different retailers and markets. Thus, ignoring assortment can create biased elasticities due to the fact that the ‘heterogeneous store’ elasticity is a function of the product’s market share only in stores that carry it, whereas the ‘homogeneous store’ elasticity is a function of the product’s average market share across all stores.

Additional modeling assumptions are needed to identify the sign of this composition bias. Tenn and Yun illustrate this by assuming away differences in weighting across retailers in order to compare elasticities between the two models. Under this assumption, if the absolute value of the price coefficient is upward biased, then the composition difference of the ‘homogeneous store’ model biases the own-price elasticities in the more-elastic direction, as seen in the following equation:

$$\varepsilon_{jj}^{\text{bias}} = (1 + \alpha^{\text{bias}}) \times \frac{1 - s_{jm}}{1 - \frac{s_{jm}}{\text{ACV}_{jm}}} - 1 \quad (4)$$

where $s_{jm} = \frac{1}{R} \sum_r s_{jmr}$ and R is the total number of retailers. The term α^{bias} denotes the percentage bias in the price coefficient, and $A_{jmr} = 0$ implies that $s_{jmr} = 0$. The upward bias on ε_{jj} is due to the fact that $\text{ACV}_{jm} \in (0,1)$ and will be further reinforced by any upward bias on α . Under the assumption that j ’s expected market share is independent of the availability of k , and again ignoring weighting differences between the models, the cross-price elasticity bias can be written as:

$$\varepsilon_{kj}^{\text{bias}} = (1 + \alpha^{\text{bias}}) \text{ACV}_{jm} - 1 \quad (5)$$

Any upward bias on α will imply that the direction of bias in the cross-price elasticities depends on the extent of availability of the relevant products, because the elasticity bias is the product of bias in the price coefficient and availability (i.e., ACV). For widely available products, the homogeneous store model will produce excessively elastic estimates, while the opposite will be true for products with sufficiently low

availability.¹¹ Finally, in addition to compositional differences, the calculation of average elasticities may also vary with the weights used across retailers. The ‘homogeneous retailer’ model allocates weights equally across markets, while the heterogeneous retailer model allocates weights equally across retailers. Given these theoretical biases, a logical next step is to examine the possible magnitudes of the bias. Tenn and Yun (2008) estimate the magnitude of the bias introduced by ignoring assortment variation through a Monte Carlo analysis, and find biases in own-price elasticity estimates of 4-18% across categories with standard deviations of 1% to 2%, and biases in cross-price elasticity estimates of -11 to -21% with standard deviations of 16% to 20%.¹²

Bias Corrections in a MNL Model

Bruno and Vilcassim (2008) presents an approach to correct the bias and compares its estimates with those of an uncorrected model. The authors begin with a standard random-coefficients logit model following the well-known model in Berry, Levinsohn, and Pakes (1995), hereafter BLP, for which utility for consumer i of good j in market m is $u_{ijm} = \mathbf{x}_{jm}\boldsymbol{\beta}_i - \alpha_i p_{jm} + \xi_{jm} + \epsilon_{ijm}$, where $\boldsymbol{\beta}_i$ is an individual-specific vector of random coefficients that represent consumer tastes for the characteristics in vector $\mathbf{x}_{jm}$. The individual-specific α_i is a random coefficient on price, and ξ_{jm} is the unobserved quality of the particular product-market pair. In order to obtain the usual logit structure for choice probabilities, ϵ_{ijm} is assumed to be distributed as a TIEV random variable. Individual preference heterogeneity is then represented as deviations from a mean, as in $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Sigma}_\beta \mathbf{v}_i$ and $\alpha_i = \alpha + \sigma_\alpha \eta_i$, where $\boldsymbol{\beta}$ and α are mean taste parameters to be estimated, $\boldsymbol{\Sigma}_\beta$ (a matrix) and σ_α (a scalar) are taste distribution parameters to be estimated, and $\mathbf{v}_i$ (a vector) and η_i (a scalar) are mean zero random variables with known distribution. Then, by defining mean utility $\delta_{jm} = \mathbf{x}_{jm}\boldsymbol{\beta} - \alpha p_{jm} + \xi_{jm}$ and consumer-specific utility $\mu_{jm}(\boldsymbol{\theta}_i) = \mathbf{x}_{jm}\boldsymbol{\Sigma}_\beta \mathbf{v}_i - \sigma_\alpha \eta_i p_{jm}$, where $\boldsymbol{\theta}_i = (\mathbf{v}_i', \eta_i')$, consumer utility can be rewritten as $u_{ijm} = \delta_{jm} + \mu_{jm}(\boldsymbol{\theta}_i)$.

If availability were known for each market, then after assuming a functional form for the distribution of the random coefficients the market shares can be written as

$$\tilde{S}_{jm}(\mathbf{a}_m) = \int \frac{a_{jm} \exp[\delta_{jm} + \mu_{jm}(\boldsymbol{\theta}_i)]}{1 + \sum_{k=1}^J a_{km} \exp[\delta_{km} + \mu_{km}(\boldsymbol{\theta}_i)]} dG(\boldsymbol{\theta}_i; \sigma_\alpha, \boldsymbol{\Sigma}_\beta) \quad (6)$$

where a_{jm} is a binary indicator of availability equal to 1 if j is available in market m and $\mathbf{a}_m$ is a vector of a_{jm} . Given a probability mass function for the availability vector, $\pi(\mathbf{a}_m)$, the model implies that observed market shares, S_{jm} , are given by

$$S_{jm} = \sum_{\text{all } \mathbf{a}_m} \tilde{S}_{jm}(\mathbf{a}_m) \pi(\mathbf{a}_m)$$

¹¹ Note that the cross-price elasticity in equation (5) does not depend, contrary to what one might expect, on the joint distribution of the two products’ availability. This feature is a consequence of the plain logit functional form of demand. Tenn and Yun point out that in more flexible demand models, such as the random coefficients logit model, the joint distribution of availability does affect the calculated average elasticity.

¹² These differences reflect differences between equation (4), where the bias is mostly driven by bias in the price coefficient when the individual market shares are small, and equation (5), for which the bias estimates vary with the availability of the product, thereby introducing substantial variation in estimates within a category.

Bruno and Vilcassim implement their method by simulating I draws from both a joint distribution of availability and an independent distribution of the taste parameters, and averaging the implied market shares across draws, before proceeding as in BLP. To construct the joint distribution of availability needed for simulation, the authors assume that the marginal availability distributions, which are observed via market-level ACV data, are mutually independent. The model also implicitly assumes that consumers' tastes do not vary systematically with availability. This may be a concern in markets for which stores choose assortments to match the tastes of nearby consumers.¹³

2.3 Demand Estimation with Observed Assortment Variation

When assortments are observed, researchers must confront the possibility that retailers select products based on unobservable components of demand. Such selection may bias demand coefficients and the estimated distribution of product-specific intercepts, thereby causing counterfactual simulations of assortments to overestimate demand for products that are not originally observed. We review a contribution to this literature, Draganska, Mazzeo, and Seim (2009), that highlights the challenges of selection on unobservables and the assumptions required for consistent estimation of a class of demand models.¹⁴ We then discuss three approaches to incorporating endogenous assortment in demand estimation.

In Draganska, Mazzeo, and Seim (2009), the utility of consumer i from purchasing product j in market m is the sum of the mean utility, δ_{jm} , a consumer-specific portion, μ_{ijm} , and an i.i.d. T1EV error term: $u_{ijm} = \delta_{jm} + \mu_{ijm} + \epsilon_{ijm}$. The term δ_{jm} includes a mean effect for price, plus fixed effects for brands and varieties.¹⁵ The term μ_{ijm} reflects the inclusion of random coefficients on price and two national brand dummies, along the lines of BLP. With a slight adjustment of notation for the random coefficients, the implied market shares can be written as a function of assortment as in equation 6 from our discussion of Bruno and Vilcassim (2008). The key difference is that Draganska, Mazzeo, and Seim observe the vector $\mathbf{a}_m$ and so they do not need to simulate its distribution. The model does not employ the usual orthogonality condition between an unobservable component of δ_{jm} (usually denoted ξ_{jm}) and a set of instruments. As the authors discuss, these unobserved components often appear in models to account for unobserved shocks to demand that are correlated with price and, in the context of assortment variation, with product-line choices. Using the unconstrained optimization approach of BLP, the authors note that they cannot separately identify mean utility levels δ_{jm} from product-market level unobservables ξ_{jm} . They also cannot treat the product-market specific intercepts as random effects because, as Berry (1994) explains, optimal price setting is inconsistent with an exogenous distribution of the relevant unobservables, conditional on product characteristics and prices. Thus, in place of the usual moments, Draganska et al. attribute the difference between observed and implied market shares to sampling error,

¹³ Changes in availability that arise from stock-out events are not well handled by this method, because such changes would be correlated with sales (i.e., stock-out events are more likely to occur during periods of high demand).

¹⁴ Their model accounts for assortment variation across markets (i.e., whether the product is offered by at least one retailer in each market), but not across retailers within a market. In this sense, the method is especially useful for, and is applied to, data aggregated to the market level, for which substitution across markets can plausibly be ruled out. Much of their paper is a contribution to supply-side modeling and counterfactual simulation, which is beyond the scope of our review.

¹⁵ We use j to denote a brand-variety pair (equivalent to a product), whereas Draganska et al. use subscripts b for brand and f for variety (flavor), separately. The distinction of the two dimensions is important for them because they also estimate a version of their model without random coefficients that includes a richer set of brand $\times$ variety fixed effects. That exercise is not essential to our discussion of identification, so we use subscript j for consistency.

and set the expectation of the squared sampling error equal to zero, thereby ruling out selection on unobservables.¹⁶

Once one acknowledges that the set of carried products is subject to selection, removing product unobservables is one of two alternatives available within unconstrained optimization approaches, such as the widely-used nested fixed-point algorithm. The second alternative is to include an explicit model of assortment selection. We now review two methods that use selection equation methods and a third approach that uses constrained optimization to incorporate selection on unobservables directly.

Iaria (2014) addresses endogeneity of assortment by first estimating a reduced-form selection equation for assortment and then incorporating the resulting inverse Mills ratio into the indirect utility function. The selection equation is given by $a_{jl} = 1(\mu_j w_{jl} + u_{jl} > 0)$, where a_{jl} is a binary indicator of product j being available in location l , μ_j are product-specific coefficients, w_{jl} are observable characteristics, and u_{jl} is an error term. Iaria notes that unaddressed selection on unobservables will always cause bias in estimated product intercepts, but the bias in the coefficients of the demand model will depend on the correlation between the observed product characteristics, including price, and the inverse Mills ratio. In this setting, the usual moment condition assumption of BLP-style models is augmented by additional identification assumptions: (i) a parametric joint distribution of the errors of the selection equation, u_{jl} , (ii) independence of the error terms in both the selection and mean utility equations from the selection equation regressors, w_{jl} , (iii) linearity of the unobserved portion of mean utility as a function of the selection error, and (iv) the exclusion of at least one element of w_{jl} from the demand equation. Given these assumptions, the adjusted indirect mean utility function appears as $\delta_{jl} = \alpha p_{jl} + \mathbf{x}_{jl}\boldsymbol{\beta} + \gamma\lambda(\hat{\mu}_j w_{jl}) + \varepsilon_{jl}$, where $\lambda(\cdot)$ is the inverse Mills ratio, and $\hat{\mu}_j$ are estimates from the first-stage selection estimation. As Iaria notes, the demand model can be used for counterfactuals in the same way as any BLP-style demand model because selection is handled prior to demand estimation.

In another selection approach, Shah, Kumar, and Zhao (Forthcoming) estimates a joint model with separate equations for (i) demand, (ii) a reduced-form assortment choice, and (iii) a reduced-form pricing choice. They use functional-form assumptions that relate observed market-level assortments and prices to observed variables and a joint distribution of the product-market error components of all three equations. The use of specific distributions for the unobserved demand and supply-side components, along with exclusion restrictions on the observable variables across the demand, assortment, and pricing models, completes the identification. They then estimate the joint likelihood function via a Markov Chain Monte Carlo procedure. In contrast to the approach of Iaria, this estimation consists of a single stage, simultaneously obtaining parameters for both demand and selection.

A third approach to selection is possible if one trades the nested fixed-point algorithm for a *constrained* optimization estimator and includes a supply-side equilibrium. This is the approach of Musalem (2015), which uses Mathematical Programming with Equilibrium Constraints (MPEC). Demand is represented by a random-coefficients logit utility model that takes the usual form, but includes, in addition to the observed components and the TIEV error term, a random effect that varies at the product-firm-market level (i.e., ξ_{jfm}). Conditional on observed characteristics and prices, the implied shares are:

¹⁶ In a model of demand for computers, Eizenberg (2014) also rules out assortment selection on unobservables but does not remove the unobservable term from demand, thereby allowing for price endogeneity.

$$s_{jfm} = \int \frac{\exp(\mathbf{X}_{jfm}\boldsymbol{\beta}_{im} - \alpha_{im}p_{jfm} + \xi_{jfm})}{1 + \sum_{f'=1}^F \sum_{k \in a_{f'm}} \exp(\mathbf{X}_{kf'm}\boldsymbol{\beta}_{im} - \alpha_{im}p_{kf'm} + \xi_{kf'm})} \phi(\boldsymbol{\theta}_{im}; \bar{\boldsymbol{\theta}}_m, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_{im}$$

where $\boldsymbol{\theta}_{im} = (\boldsymbol{\beta}'_{im}, \alpha_{im})'$, $\phi(\cdot)$ is the density function of the vector of random coefficients, and $(\bar{\boldsymbol{\theta}}_m, \boldsymbol{\Sigma})$ are the parameters of that density function. The set of products offered by firm f in market m is denoted a_{fm} . The unobserved ξ_{jfm} are assumed to be distributed normally with mean 0 and variance φ_{ξ}^2 , which is estimated.

The supply-side of the model includes a two-stage game, in which brands first select assortment according to a Bayesian-Nash equilibrium, and then select prices according to Bertrand-Nash. In the first stage, brands “form beliefs about the profits of their competitors and anticipate the prices and demand they would observe for any given set of products that could be introduced in the market.” While we do not discuss the supply-side model in detail here, it is fundamental to the estimation approach, in that the demand model cannot be estimated separately from the supply model. The final, constrained, optimization problem is a maximum-likelihood problem with three constraints: (i) the observed market shares must match the implied market shares, (ii) firms’ first order conditions from price-setting must hold, and (iii) firms’ beliefs about the likelihood of their competitors’ choice of assortments must be consistent with the model. Firms’ beliefs are modeled as a function of the common knowledge components of firm profits, including all ξ s within the market, and a T1EV private information component, thereby generating a logit form. The assumed distribution of ξ enters the likelihood function, and ξ enters all three of the constraints: the first two constraints through the share equation and the third constraint because ξ affects the profitability of a particular assortment.

Musalem recovers a distribution of ξ for the population of products, not just those that were included in the assortment, and uses that distribution to simulate demand under counterfactual assortments. This capability is not without potential costs. The model requires an *a priori* marginal distribution for ξ .¹⁷ Additionally, since estimation of the model requires the supply side, incorrect assumptions on the supply side (such as the information, time, and choice structure of the game, or the assumed form of marginal and fixed costs) can affect the demand estimates.

In summary, an applied researcher interested in demand and assortment has a choice between (i) assuming away selection on unobservables, (ii) incorporating a reduced-form selection equation in a prior stage, (iii) incorporating a reduced-form selection equation and correlation of errors across equations, and (iv) modeling selection on unobservables through equilibrium conditions with a structural supply-side. Table 1 relates the basic frameworks of the latter three approaches. Iaria (2014) and Shah, Kumar, and Zhao (Forthcoming) will be sensitive to misspecification of the supply equation(s) and the distributions of unobservables, while Musalem (2015) ties the consistency of the demand estimates to the validity of the chosen supply-side and equilibrium. When choosing whether to endogenize assortment through any of these methods, one must address the question, “How serious are the consequences of assuming away selection on unobservables?” The answer depends to a significant extent on the definition of market, in terms of geography, time, and products; the diversity of the set of locations; and the extent to which one can capture determinants of demand using observed variables and fixed effects. Two applications explore

¹⁷ Musalem refers to Jiang, Manchanda, and Rossi (2009) as evidence that misspecification of this distribution may not have a substantial impact on demand estimates.

this question with detailed data in particular markets: the national market for shoes, and the market for margarine in one metropolitan area.¹⁸

Table 1: Models for Observed Assortment Data with Selection on Unobservables

	Models		
	Iaria (2014)	Shah, Kumar, and Zhao (Forthcoming)	Musalem (2015)
Estimation of Demand and Supply	Separately	Jointly	Jointly
Assortment Selection on Unobservables	Reduced-form estimation prior to demand estimation	Reduced-form simultaneous to demand estimation	Structural
Equilibrium Assumptions	Optional	None	Necessary

In an application to the national shoe market Quan and Williams (2014) uses sales data from an online retailer and develops an estimation approach that combines national sales data with local sales levels but not local characteristics. They include a product-location random effect with product-specific variance to account for local tastes, and find that local variation in demand plays an important role in determining optimal assortments when local assortment size is restricted (e.g., because of stocking costs). The second application, Chintagunta, Dubé, and Goh (2005), investigates the importance of unobserved components of demand for margarine sales in Denver, CO. The authors use household-level panel data to compare random-coefficient models estimated (via ML) with and without an error term that is the product-time analogue of the product-location unobservable in the assortment models. Their results, which do not address variation in assortment *per se*, suggest that estimates of price coefficients and their variances are significantly impacted by the inclusion of such unobservables for the market they study.¹⁹

3 Stockout Events

Both assortment choice and stockout events can have substantial effects on firm profits, and firms in many industries identify inventory management as a critical strategic decision. When a product stocks out, observed purchases no longer reflect the true underlying demand, either for the stocked-out good or for its available substitutes. Failing to account for stockouts, especially in markets for which products are perishable, seasonal, or have storage costs, may lead to biased estimates of demand.²⁰ On the other hand, stockouts many introduce a helpful source of exogenous variation in the data. In this sense, stockouts provide researchers with both an important obligation (to avoid biased estimates of demand) and a valuable opportunity (to leverage exogenous variation in choice sets across consumers).

¹⁸ Bronnenberg, Dhar, and Dubé (2007) also document variation in product tastes across regions.

¹⁹ The endogeneity of observed characteristics other than price is discussed in Crawford (2012).

²⁰ An additional challenge, even for models that adjust for availability sets, arises when the distribution of consumer tastes and the (multinomial) distribution of stockouts is not independent. Even if consumers do not directly sort themselves according to stockout probabilities, they may sort themselves according to tastes (e.g., by shopping on different days of the week or at different times of day) that correlate with stockout probabilities. Scanner panel data with observed demographics and stockout incidences might be used to partially correct for this problem.

Boundaries

Stockouts of consumer goods in retail settings do not fully mimic “sell-outs” for other types of products, such as tickets to a music concert, sports event, or plane flights. For capacity-constrained perishable goods that can be sold well in advance, such as event tickets, demand can be a function of time until the expiration of the good (e.g., Escobari (2012) and Lazarev (2013)). Such products can also become available all at once, which can, in combination with evident underpricing in some markets, generate competition among consumers to purchase before capacity constraints bind in a primary market (e.g., Leslie and Sorensen (2014) and Sweeting (2012)). Several approaches to varying availability can be found in studies of these markets. Some authors with access to individual sales transaction data use a static demand framework and assume availability is fully observed for all purchasing customers (e.g., (Talluri and van Ryzin (2004) and Vulcano, van Ryzin, and Chaar (2010)). Others who are interested in primary and secondary market for tickets have modeled (i) purchase with arrival sequence-dependent availability sets (e.g., Leslie (2004)), and (ii) a secondary market among consumers with rational expectations in equilibrium (e.g., Leslie and Sorensen (2014)). Researchers working with daily data on sales and availability have assumed that, on days during which capacity for a particular good is exhausted, consumers who preferred one product are randomly selected into either obtaining the good or being forced to substitute to the outside good (e.g., Williams (2013)). Others, working with quarterly market data, have addressed unobserved availability with product fixed-effects (e.g., Berry and Jia (2010)). As these concerns and approaches are not typical in retail settings, we do not address them in greater detail.

On the supply side, stockouts can affect retailers’ decisions to offer price reductions (e.g., Aguirregabiria (1999)) and may impact competition between upstream manufacturers (e.g., Aguirregabiria (2005)).²¹ Finally, the OR and RM fields have developed a theoretical literature of optimal firm choices in the face of demand with stockouts (see, e.g., Tan and Karabati (2013) and Stavroulaki (2011)). Before reviewing methods for estimating demand in the presence of stockouts, we first describe empirical evidence on the prevalence of stockouts and the consequence of ignoring them.

3.1 Prevalence and Sources of Variation

To understand the extent and consequences of retail stockouts, Gruen, Corsten, and Bharadwaj (2002) collects and analyzes 40 studies of stockouts of grocery and drugstore items, which vary by categories considered and geographic region. While acknowledging that there may be variation in the definition of and calculation method for stockouts across the studies, they report that the average stockout rate, usually defined as the percentage of SKUs that are out-of-stock at a given moment, across all categories, is 8.3%.²² Among the six product categories for which Gruen et al. present averages, the average stockout rate varied from a low of 5.3% for salty snacks, to a high of 9.8% for hair care products. In the studies they examined that reported stockout rates by day of the week, the rate was highest on Sunday (an average 10.9% rate), and then generally declined over the course of the week to end with a low of 7.3% on Saturday. Based on a single study that considered the duration of stockouts, Gruen et al. report that

²¹ Both of these papers use the same underlying data source, which describes inventory orders by outlets to a central distribution warehouse of a grocery chain in Spain. Although Aguirregabiria (2005) develops a model that accounts for stockouts when estimating demand, the data are unusual for this literature because they are aggregated at the monthly level. In the model, a consumer forms an expected utility of the optimal choice at a store by integrating out the probability of each possible availability set (conditional on stockout probabilities at the store) over the expected maximum utility within the store.

²² This is essentially the same as the rate found several years earlier by Andersen Consulting (1996).

20% of stockouts last less than eight hours, 25% last between 8 hours and one day, 36% last between 1 day and 3 days, and only 19% last for three days or more. The reviewed studies also gave estimates of sales loss due to stockouts, and the average loss across all relevant studies was 3.9% of sales.

Using perhaps the largest overall survey of stockouts in the grocery channel, Matsa (2011) calculates stockout rates from Bureau of Labor Statistics (BLS) microdata on supermarket products used to compute the CPI. His data contain more than four million observations on in-stock status of over 200 000 products in over 10 000 stores in the U.S. He reports that the mean stockout rate across all categories is 4.3%, and attributes the difference between this estimate and the more typical estimates of 8% to two sources. First, CPI data are often collected during the day on weekdays, when stockout frequencies are lower. Second, the CPI does not count a product as out-of-stock if a store employee states that the product will be replenished the same day.

Several individual industry demand estimation papers that account for stockouts have presented evidence of the frequency of stockouts in their datasets. Aguirregabiria (2005) averages stockout probabilities for 15 different categories of products using inventory data from a supermarket distribution center. Salt has the lowest observed stockout probability, at 1.2%, and the highest rate occurs for razors, at 9.6%, while the mean, across the categories, is 5.6%. The stockout rate variation within categories (across brands) is much greater than the variation across categories. In the olive oil category, for example, brand-specific stockout probabilities range from 3.4% to 11.7%, while for Tuna the range extends from 2% to 20%. In their study of stockouts in the vending machine industry, using data from 54 machines on the campus of Arizona State University, Conlon and Mortimer (2013) estimate that most products are out-of-stock during two to three percent of the time, while a few products, especially pastries and pretzels, are out-of-stock as often as about 28% of the time. Musalem et al. (2010) present out-of-stock incidence rates, as a percentage of days in which a stockout occurred, for shampoo products from several supermarkets in Spain, and found out-of-stock rates of 0% to 42%, with a median of 17%.

Campo, Gijsbrechts, and Nisol (2003) present discrete bins of product stockout rates and market shares, from both the cereal and margarine categories of a European retail chain, that suggest that products with large market shares stockout less frequently than products with small market shares. Che, Chen, and Chen (2012), on the other hand, find large positive correlations (between .73 and .99) between market shares and average number of stockouts at the brand level, based on two years of scanner data from the grocery channel, in six product categories. The difference in results could be due to the much shorter dataset used in Campo et al. (28 weeks compared with two years) or due to the fact that Che et al. consider brands rather than products in the calculation of the correlations.²³ The scanner data of Che et al. also show that product prices are significantly lower during days that feature stockouts than during other days, suggesting that price coefficient estimates will be biased in models that do not account for stockouts.

A number of studies from the marketing, retail, and psychology literatures have investigated categories of consumer response to stockouts through both experiments and surveys. The categories of behavior are typically divided amongst (i) canceling purchase, (ii) delaying purchase, (iii) switching products, and (iv)

²³ The difference could also come from the fact that Che et al. chose a location with only one large supermarket to avoid issues of competition among retailers. Retailers who face greater competition could have different optimal inventory policies, leading to different correlations between market share and stockout frequency. See Matsa (2011) for an examination of stockouts and retailer competition.

switching stores. Evidence of the extent of each of these categories of behavior is relevant for researchers seeking to estimate demand models. Purchase cancellation is already accounted for by most models of demand, but experimental evidence of its frequency could be useful for considering the probability of substitution to the no-purchase option. Purchase delay suggests that consumers make dynamic decisions about purchase. The ability to delay is often not incorporated into demand models for stockouts, in part because researchers rarely observe household stockpiles/inventories. Evidence of extensive product switching in response to stockouts bolsters the argument for accounting for availability variation in demand models. Finally, the degree to which store switching may bias demand estimates depends on the research question and the context of the model. For example, when studying demand at a single retail location, store switching may be treated as equivalent to purchase cancellation. On the other hand, store switching may be problematic for market-level demand models because, without observing or modeling store choice, such models implicitly assume that store switching and search are costless.²⁴

Summarizing the evidence from their review of 40 studies, Gruen, Corsten, and Bharadwaj (2002) report the average, across 11 categories of supermarket products in the U.S., proportions of consumer response types. A significant portion of consumers, 40%, substitute to a different product within the store, 32% substitute to a different store, 17% delay purchase, and 11% cancel purchase. In addition, response shares vary significantly across product categories.²⁵

3.2 Consequences of Ignoring Stockouts

Stockouts can bias several estimates of interest in demand models, including own- and cross- price elasticities, demand levels under counterfactual policies (e.g. changes to price), and coefficients on price and other product characteristics (both the mean, and other distributional parameters in the case of random coefficients). The effect of an unobserved stockout on estimated demand levels is fairly straightforward; demand for products that are unavailable during some periods will be underestimated due to *censoring*, and demand for other products, particularly close substitutes, will be overestimated due to *forced substitution*. It is possible, however, for these two effects to interact. A product may become out-of-stock only as a result of forced substitution from a close substitute that has already stocked-out.

The effects of stockouts on estimated demand parameters and elasticities can vary. When stockouts occur during promotional price cuts, consumers may appear less price sensitive than they are in reality, due to demand censoring. For example, Che, Chen, and Chen (2012), using supermarket data, finds that accounting for stockouts leads to an increase of between 4% and 21% in the absolute value of mean price coefficients, depending on the product category. In other work on supermarket categories, Aguirregabiria (2005) also finds that estimated price sensitivity increases (by 20%) after accounting for stockouts. In contrast, a study of shampoo sales from several stores of a supermarket retailer by Musalem et al. (2010) shows decreases of 10.5% in the absolute value of the mean (random) price coefficient and 28.6% in its average variance.²⁶

²⁴ The cost of switching and search almost certainly vary by product and retail sector. Rhee and Bell (2002) provide evidence that many grocery customers only rarely switch the store of their largest weekly purchase from week to week.

²⁵ As some of the research on types of consumer response to stockouts is conducted via surveys, one should note that, at least in some laboratory settings, consumers' responses can be influenced by whether or not they expect to justify their choices (see Kramer and Carroll (2009)).

²⁶ Authors' calculations based on table 4 of Musalem et al. (2010).

Price elasticities in the logit model (see equations (2) and (3) in the previous section) can be biased by stockouts both because price coefficients are biased, and because the market shares are distorted through censoring and forced substitution. Since the net bias is the product of two biases, for some products and some biases of the price coefficient the elasticity bias can be given a particular sign, while for others the sign is not determined without knowing the relative magnitudes.

Estimates of coefficients of observed product characteristics other than price can also be biased by unobserved or ignored stockouts. Musalem et al. (2010) reports changes of between 9.8% and 73.3% in the magnitude of estimated brand fixed-effects when moving from a model that ignores stockouts to their model that addresses the partially-unobserved stockouts found in their data.²⁷ Conlon and Mortimer (2013) finds significant increases in the parameter estimates of a nested-logit model when moving from a method that ignores stockouts to a method they develop for dealing with partially-unobserved stockouts.

Where possible, researchers may be tempted to simply remove periods containing stockouts from the data they use for estimation. This choice has consequences for estimation because stockouts may happen more frequently when unobserved demand shocks are high. For example, Conlon and Mortimer (2013) finds that average sales during four-hour periods with stockouts are about 2.5 times as large as during periods without stockouts, and Che, Chen, and Chen (2012) finds that stockouts are especially likely to occur during promotional price cuts. Thus, discarding stockout periods may select on demand shocks and price (as well as reduce the observed price variation in the data).

Alternatively, one could approximate availability sets by classifying each product as either in-stock or out-of-stock for the entirety of each period.²⁸ Availability would then vary across, but not within, periods. This classification requires one to distinguish between periods with stockouts and those with positive inventory but zero sales. Some researchers have followed this approach by first estimating confidence intervals or setting thresholds for normal sales, then defining periods with exceptionally low sales as containing the beginning of a stockout, and finally classifying the given product as out-of-stock during subsequent zero-sales periods (see Campo, Gijsbrechts, and Nisol (2003) and Karabati, Tan, and Öztürk (2009)).²⁹ The drawback of “always or never” classification is that, as Conlon and Mortimer (2013) demonstrate, it contradicts the assumption of multinomial choice that underpins models of discrete-choice demand. This contradiction arises because the distribution of observed sales across availability sets that is consistent with observed stockouts is already determined, without any free parameters, by the parameterization of the discrete-choice model. In the next section we review methods for incorporating stockouts into the demand model itself.

3.3 Demand Models for Partially Unobserved Stockouts

Anupindi, Dada, and Gupta (1998) presents one of the earliest attempts to estimate discrete-choice demand when availability varies and is less than fully observable due to stockouts. In their model, customers who prefer each product j arrive in each period according to an independent Poisson distribution. When a product j is out of stock, a set of probabilities captures the likelihood of substitution

²⁷ Authors' calculations based on table 4 of Musalem et al. (2010).

²⁸ Periods during which a stockout begins will be classified as having in-stock availability to be consistent with observed positive sales.

²⁹ These methods may be infeasible for slow-moving products because any positive threshold for identifying exceptionally low sales would misclassify a significant portion of periods during which the product was available but not purchased.

to each of the other options. If the second-choice item is also out of stock, the customer is assumed to choose the outside good. This *one-stage substitution* assumption restricts the number of parameters that must be estimated in this essentially non-parametric model of substitution, but also places strong restrictions on substitution patterns. Without restricting substitution to a single-stage, the number of parameters to be estimated would grow much too quickly in the number of products to permit feasible estimation. Note that the probabilities are not projected onto a set of observed characteristics, rather they are simply product-space likelihoods. To address stockouts that happen in the middle of a period, Anupindi et al. use the EM algorithm, and treat as missing data the total time for each availability regime (which is important because of the assumptions on customer arrivals) and the sales of each product under each availability regime.

Since Anupindi, Dada, and Gupta (1998) the development of models for unobserved stockouts has largely but not exclusively coincided with the MNL tradition, while models for observed stockouts have often been developed without logit or even RUM approaches, most likely because the methods emerged in different fields with access to different data structures.³⁰ We begin with the former category and then briefly discuss the latter.

Conlon and Mortimer (2013) and Musalem et al. (2010) develop demand estimation approaches for use with periodic store data in which one observes either levels of inventory or indicators of inventory availability at the end of each period.³¹ The periodicity of the data is an important feature of both models, because it implies that at least some stockout events are unobserved. Thus, the researcher must estimate either the timing of the stockout event, or a distribution for sales of other products across the different availability sets that arise from the stockout event.

The key insight of the Conlon and Mortimer model is that discrete-choice models of demand, for which sales follow a multinomial distribution, imply a distribution for the sales of each product j across different availability regimes when changes in availability result from a stockout of another product, k . To illustrate, let the probability that an individual i chooses product j during period t be denoted p_{ijt} and consider the example of a single stockout event. Denote the number of consumers who arrive before k stocks out but choose a product other than k , as r_t . The multinomial distribution implies that r_t follows a negative binomial distribution, for which the probability mass function (p.m.f.) is given by:

$$f(r_t, y_{kt}, p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta})) = \frac{(y_{kt} + r_t - 1)!}{r_t! (y_{kt} - 1)!} p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta})^{y_{kt}} (1 - p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta}))^{r_t}$$

in which p_{ijt} is specified as $p_j(\mathbf{a}_t, \mathbf{x}_{jt}, \boldsymbol{\theta})$, $\mathbf{a}_t$ denotes the availability set during period t including product k , $\mathbf{x}_{jt}$ is a set of observable characteristics of the choice scenario, y_{kt} denotes sales of product k in period t (before k stocks out), and $\boldsymbol{\theta}$ is a vector of parameters to be estimated. Assuming a market size of M_t consumers in period t implies that the conditional p.m.f. is given by:

³⁰ The economics and marketing fields have focused on periodic inventory data structures, and the revenue management and operations research fields, with access to individual transaction data or perpetual inventory data, have developed models focused on observed stockouts.

³¹ Chen and Mersereau (2014) develop another method that is helpful for retailers but unusual for researchers, because the agent that estimates demand also controls inventory levels in a dynamic setting. The agent faces a tradeoff between greater storage costs of extra items and learning about true demand levels by ordering enough inventory to avoid stockouts.

$$h(r_t; y_{kt}, p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta}), M_t) = \frac{f(r_t; y_{kt}, p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta}))}{F(M_t; y_{kt}, p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta}))}$$

By integrating over this distribution, one can calculate the expected sales of each product j before k stocks out, conditional on each product's observed total sales in the period and the parameter vector $\boldsymbol{\theta}$, given by:

$$\begin{aligned} & E[y_{jt}^{a_t} | y_{jt}; \boldsymbol{\theta}] \\ &= y_{jt} \sum_{r_t=0}^{M_t-y_{kt}} \frac{r_t \cdot p_j(\mathbf{a}_t, \mathbf{x}_{jt}; \boldsymbol{\theta})}{r_t \cdot p_j(\mathbf{a}_t, \mathbf{x}_{jt}; \boldsymbol{\theta}) + (M_t - y_{kt} - r_t) \cdot p_j(\mathbf{a}_t', \mathbf{x}_{jt}; \boldsymbol{\theta})} h(r_t; y_{kt}, p_k(\mathbf{a}_t, \mathbf{x}_{kt}; \boldsymbol{\theta}), M_t) \end{aligned}$$

in which $\mathbf{a}_t'$ denotes the availability set after k stocks out. Using this expectation, one can write an expected likelihood function. Conlon and Mortimer propose the use of the EM algorithm for estimation, because direct maximization of the expected log-likelihood is computationally burdensome.

The approach leverages the fact that the sufficient statistic for estimating the multinomial distribution is total sales within a period (and not an ordered sequence of sales), and it has two advantages. First, one need not consider sales of products other than j and k when computing the expected sales of j . Second, the expectation is a single summation, so one does not need to keep track of every possible value of $y_{jt}^{a_t}$ when maximizing the likelihood. Conlon and Mortimer demonstrate the application of their method using vending machine data.³² For their application, the authors employ two different specifications for the parametric form of $p_j(\cdot)$: a random-coefficients logit, and a nested logit.

A potential limitation of the approach of Conlon and Mortimer is that, as the number of simultaneous stockouts increases, the computational burden of the model grows quickly. Intuitively, this is because simultaneous stockouts increase the number of unique availability regimes within a time period, and thus increase the amount of missing data that must be estimated. Note, however, that the computational burden does not increase when the data contain large sales volumes, because estimation requires only expected sales in each availability regime.

Musalem et al. (2010) provides a second modeling approach to account for stockout-induced unobserved availability variation. Instead of treating the summary statistics of sales across availability regimes as missing data, they treat the entire sequence of sales and availability sets as missing data. Their model assumes that, for each period, sales and end-of-period availability status are observed. These sales and availability data are potentially consistent with many different sequences of individual sales and availability sets. The authors write the likelihood function of a random-coefficients logit model (with random product-market effects) conditional on the unobserved, augmented data, and then sum the likelihood over the set of sales and availability set sequences that are consistent with the observed data. The total likelihood is then

³² In their application, stockouts are a regular occurrence, but price variation is very limited. Due to the lack of price variation, Conlon and Mortimer do not estimate a price coefficient. Note, however, that their method does allow for estimation of price coefficients.

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\xi|S,I}) = \prod_{m=1}^M \prod_{t=1}^T \sum_{(\mathbf{W}_{tm}, \mathbf{A}_{tm}) \in \Omega_{tm}(\mathbf{S}_{tm}, \mathbf{I}_{tm})} L_{tm}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_{\xi} | \mathbf{W}_{tm}, \mathbf{A}_{tm})$$

in which $\mathbf{W}_{tm} = (w_{1tm}, \dots, w_{N_M tm})$ is a vector of individual choice indicators, $\mathbf{A}_{tm} = (a_{1tm}, \dots, a_{N_M tm})$ is a vector of unobserved product availability indicators, $\mathbf{S}_{tm}$ is a vector of product-level sales for each store period, and $\mathbf{I}_{tm}$ is a vector of initial inventory levels.

As in the Conlon and Mortimer model, this total likelihood is too computationally burdensome to be feasibly estimated, for two reasons. First, it requires summation across all sequences of sales and availability that are consistent with the observed sales and inventory, and the set of consistent sequences may be large. Second, the random-coefficients portion of the model requires simulation of the integrals because they lack closed form. The solution that Musalem et al. devise is to treat the missing data as parameters to be estimated and use Markov Chain Monte Carlo methods, in a Bayesian framework, to sample from the full distribution of both missing data and demand parameters. In order to lighten the computational burden, they use a Gibbs sampling mechanism, which requires conditional, rather than full, distributions, after randomly partitioning the (augmented) sequences of sales into pairs in order to sample from the possible consistent sequences of $\mathbf{A}$ and $\mathbf{W}$.

In contrast to the method of Conlon and Mortimer (2013), this approach does not become computationally more difficult when the number of simultaneous stockouts increases. Simultaneous stockouts will affect the sets of $(\mathbf{A}, \mathbf{W})$ pairs that are consistent with observed data, but the computational burden of the Gibbs sampling scheme will not increase. On the other hand, larger sales volumes will increase computational demands, because the method samples from the entire sequence of sales and availability sets, which will slow the mixing time of the Markov chain. In other words, larger sales volumes increase the number of pairs of $(\mathbf{A}, \mathbf{W})$ that are consistent with $\Omega(\mathbf{S}, \mathbf{I})$, and sampling from this larger space requires more iterations of the Markov chain. In other words, larger sales volumes increase the number of pairs of $(\mathbf{A}, \mathbf{W})$ that are consistent with $\Omega(\mathbf{S}, \mathbf{I})$, and sampling from this larger space requires more iterations of the Markov chain. Thus, this method can be computed most quickly in datasets with low per period sales volumes. For example, in the authors' application to shampoo sold in several supermarkets, average daily unit sales of most products are lower than 0.3, and the sum of average daily sales across all products is less than five.

Table 2: Stockout Demand Estimation Comparison

	Conlon and Mortimer (2013)	Musalem et al. (2010)
Required Data	Sales, starting and ending inventory levels	Sales, starting and ending inventory indicators
Strength	Handles large sales volumes	Handles many simultaneous stockouts
Limitation	Computational burden increases with number of simultaneous stockouts	Computational burden increases with sales volumes

Table 2 summarizes some of the main features of both approaches. The definition of periods matters for the implementation of both models. The ability to handle low sales volumes is an advantage of the likelihood approach that both models adopt. More commonly used demand estimation methods (e.g., BLP) often assume large sales volumes in order to invert market shares and construct moments for

estimation via Generalized Method of Moments (GMM) or Method of Simulated Moments (MSM).³³ This approach does not work well when sales levels are low, because market shares will be measured with error in these cases, which violates a condition necessary for the market-share inversion.³⁴ On the other hand, if sales volumes are high and many products stockout simultaneously, both approaches will be computationally challenging.

One possible solution in that situation is to shorten the definition of a period, and thus limit the number of sales and the number of simultaneous stockouts within each period. Of course, that is possible only when one can reliably disaggregate the data from the existing periods. Furthermore, the data should include either inventory levels or inventory indicators for all products at each replenishment occurrence. When product replenishment cycles vary across products, retailer data may not contain accurate inventory records at the beginning and end of each sales period as defined by the replenishments.³⁵

These demand estimation methods require inventory information at the beginning and end of each period. The consistency of those estimators could be threatened by endogeneity or inaccuracy of the inventory records. As Musalem et al. (2010) explain, inventory could be endogenous either because local retailers make optimal inventory decisions in response to local demand, or because serially-correlated demand shocks induce correlations between current demand and current inventory through the effect of prior demand on current inventory. Matsa (2011) presents evidence that inventory controlled by retailers (e.g., through replenishment from a regional distribution center for a chain) responds differently to changes in local competition than inventory controlled by manufacturers, for products within the same retailer. This suggests that researchers should be concerned about endogenous inventories resulting from strategic retailer behavior. Musalem et al. mention two methods for addressing this possible endogeneity within their demand estimation approach. First, one could use a Bayesian instrumental variables approach as explained by Yang, Chen, and Allenby (2003), in which an equation with inventory on the right-hand side, at least some excluded (from the demand model) instruments on the left-hand side, and an error term that is allowed to be correlated with unobserved demand shocks is included in the likelihood model. The second method follows Manchanda, Rossi, and Chintagunta (2004), and involves replacement of the excluded instruments of the previous method (if not available) with the mean of the demand shocks and the assumption that the error term of the inventory model is uncorrelated with the deviation of the demand shock from its mean.

When demand estimation requires inventory levels or indicators, inventory record inaccuracy may lead to incorrect inference. Inventory records may be incorrect because of incorrect sales records (e.g., selling organic produce as regular produce), misplaced items within the store, items lost in the stockroom, employee or customer theft, failure to update records during returns and exchanges, and errors by the distribution center. Raman, DeHoratius, and Ton (2001) finds that, for a large unnamed retail chain, more than 65% of inventory records for store-SKU pairs do not match physical inventory levels, and the

³³ Conlon and Mortimer (2013) provide a discussion of the use of GMM-based estimation, which may be used in their model instead of likelihood-based approaches, when sales volumes are large (so that market shares are measured without error). The use of GMM in estimation is handled by the Generalized EM algorithm of Dempster, Laird, and Rubin (1977).

³⁴ Indeed, a small literature on dealing with zero-sales observations has developed. See, for example, Gandhi, Lu, and Shi (2014), Gandhi, Lu, and Shi (2013), and Briesch, Dillon, and Blattberg (2008).

³⁵ Heterogeneous period lengths also necessitate varying market sizes, because failing to adjust market size will produce large swings in the share of the outside good.

absolute difference between the records and the physical levels averaged 35% of target inventory levels. Lee and Özer (2007) reviews the production and operations management literature on models and heuristics for optimal retailer policies (e.g., replenishment and auditing) in the face of several sources of inventory record inaccuracy. DeHoratius, Mersereau, and Schrage (2008) presents a method for estimating true inventory levels with Bayesian updating.

Within the RM and OR fields, which primarily work with observed stockouts, the relevant methods are often referred to as “demand unconstraining” or “demand uncensoring.” Many of the unconstraining methods were developed for estimation of single-product demand (see, for example, Besbes and Muharremoglu (2013) and Li, Lim, and Wang (2014)), and hence do not incorporate forced substitution to other goods. Of the multi-product methods that allow for stockouts, most assume that availability is fully observed for all customers that purchase a good (e.g., Talluri and van Ryzin (2004), Kök and Fisher (2007), Karabati, Tan, and Öztürk (2009), Vulcano, van Ryzin, and Ratliff (2012), Newman et al. (2014), and van Ryzin and Vulcano (2014)), likely because these researchers often work with point-of-sales or booking data for which individual transactions and inventory levels are observed. The availability set is usually not known, however, for customers who choose not to purchase. To address this unknown input, some of the contributions use the EM algorithm (e.g., Talluri and van Ryzin (2004) and Vulcano, van Ryzin, and Chahr (2010)). Others have used the EM algorithm to address that fact that sales data contain only observed choices, not first and second choices, for each customer (e.g., Kök and Fisher (2007)). Since most of these models that permit stockout-induced substitution do not begin with random utility maximization, they must choose another structure for the substitution process, which is sometimes referred to as an exogenous, rather than utility-based, model of substitution.³⁶ For example, Smith and Agrawal (2000) uses three different substitution structures: (i) market-share based substitution, which is similar to multinomial logit without nests or random-coefficients, (ii) adjacent substitution, in which all products are ordered and substitution occurs equally and only to adjacent products, and (iii) one-item substitution, in which all consumers are assumed to substitute to a single product regardless of the product from which they substitute. Guo, Xiao, and Li (2012) reviews the literature on demand unconstraining in RM and categorizes the methods according to the number of products they consider (one or multiple), the industry for which they were developed or to which they were applied, and the basic approach they employ. Azadeh, Marcotte, and Savard (2014) provide another taxonomy of uncensoring methods in RM.

3.4 Stockout-derived Variables in Demand Models

Some demand estimation methods designed for observed stockouts include stockout-derived variables in the index function of the multinomial choice probability. These additional variables capture both historical considerations (e.g., stockouts during the previous shopping trip) and contemporaneous contextual features of the choice setting (e.g., stockouts of products with similar characteristics). The inclusion of these stockout-derived variables can relax the usual independence of irrelevant alternatives (IIA) property of multinomial logit models, at least during stockouts.³⁷ For example, Campo, Gijsbrechts, and Nisol (2003) includes $\sum_A \rho_A OOS_{A,jt} + \phi OOS_{j,t-1}$, where t represents a particular shopping trip, A is an index for discrete product characteristics, $OOS_{A,jt}$ is the number of out-of-stock products that are

³⁶ Notable exceptions include Vulcano, van Ryzin, and Chahr (2010), Vulcano, van Ryzin, and Ratliff (2012), and Newman et al. (2014).

³⁷ Some stockout-relevant demand models from marketing and retail also include purchase incidence or quantity choice decisions and include stockout-derived variables in the functional forms of these decisions as well (see Campo, Gijsbrechts, and Nisol (2003) and Breugelmans, Campo, and Gijsbrechts (2006)).

similar to j in characteristic A , and $OOS_{j,t-1}$ is an indicator of the in-stock status of j in the previous shopping trip. For positive values of the ρ parameter, a product j gains (loses) market share, relative to another product i , when it is more (less) similar to the out-of-stock product.

Che, Chen, and Chen (2012) observes that since stockouts are likely correlated with unobserved demand shocks, stockout-derived variables are endogenous. They address the endogeneity through a control function approach Petrin and Train (2010) and find that correcting for endogeneity results in price and OOS coefficients (similar to those in Campo, Gijsbrechts, and Nisol (2003)) that are larger in absolute value.³⁸

Evidence for the importance and direction of these OOS-derived variables is mixed. Che et al. report that the coefficient of their stockout similarity variable is positive and significant at the 5% level across all six categories of products for which they estimate demand, suggesting that stockouts lead to greater substitution to similar products than would be predicted by an MNL model without an OOS variable in the utility function. They also find a negative and significant coefficient across five categories of products for the interaction between similarity and purchase shares, meaning that the positive effect of stockout similarity is counteracted when the cumulative purchase share of the OOS product is higher. For Campo et al., whose latent class approach generates distinct parameters values for a number of discrete classes chosen according to goodness of fit statistics, the statistical significance of OOS similarity variables varies from highly significant to insignificant across the two markets (margarine and cereal) and across the (latent) classes of consumers.³⁹ The authors' preferred margarine estimation has only a single class, and for that class the brand variable is significant at the 1% level with a positive sign, implying increased substitution toward similar products, while for cereal the estimate of the brand coefficient, for the only class for which the brand coefficient was statistically significant, has a negative sign.⁴⁰ Finally, by comparing a model that ignores stockouts entirely with a second that adjusts availability sets and a third that both adjusts availability sets and includes OOS-derived variables, Che et al. provide evidence that, on average, the effect of adjusting availability sets is four times as large (14.5% compared to 3.4% change in the price coefficient) as the additional effect of including stockout-derived variables. In summary, stockout-derived variables may be useful in modeling consumer demand, but they may matter for only subsets of consumers, only some goods (perhaps depending on consumer purchase frequency within the category), they are likely endogenous, and they appear to be less important than simply accounting for stockout variation in the availability set.

3.5 Experimentally-Induced Stockouts

Thus far the estimates of bias that we have summarized are simply differences between estimates from models that ignore stockouts and those that do not. Such comparisons do not necessarily capture the true bias because both models could be biased (e.g., for reasons other than stockouts). Conlon and Mortimer (2010) provides a rare chance to compare predicted substitution from two frequently-used demand models

³⁸ The key assumptions for the control function approach are that the endogenous variables can be written as auxiliary functions of observables; both those already included in the logit index function, denoted x , and others (i.e., excluded instruments), denoted z , such that x and z are independent of both the error term of the index function (i.e., ξ_{jt}), and the error term of the auxiliary function. See Petrin and Train (2010) for details.

³⁹ Note that their estimates depend on other portions of the model not discussed here, including a purchase incidence and purchase quantity portion of the total likelihood.

⁴⁰ See Breugelmans, Campo, and Gijsbrechts (2006) for additional mixed evidence in a similar model.

with substitution inferred from exogenously-induced stockouts via a field experiment. The authors collaborated with a vending machine operator to remove products with large market shares from vending machines, and compare the sales of confections during these ‘treatment’ weeks with sales during control weeks, in which no products are removed. They estimate both a nested-logit and a random-coefficients logit demand model using data from both control and treatment weeks. They find that both models overestimate substitution to the outside good, underestimate sales of top substitutes, and overestimate sales of other goods, relative to the non-parametric treatment-effects estimator. The authors propose four possible reasons for the inaccuracy of the structural demand estimates. First, the process of matching treatment and control weeks, which is used to generate non-parametric estimates of the treatment effect of the stockout, is noisy due to changes in product mix within vending machines, and unobserved exogenous changes in location-specific demand. Second, the parametric functional forms may be mis-specified (e.g., due to the logit error term). A third potential difficulty for the models is the endogeneity of assortment within vending machines, because variation in assortment across machines contributes to identification of the models. As the authors explain, substitution between two products that are always observed together in the data is identified in part by changes in shares of other products that vary according to endogenous assortment decisions across vending machines. The fourth proposed explanation is the absence of price variation for identification of substitution.⁴¹

Experimental approaches are not a panacea for demand estimation, however, precisely because they do not impose structure that may restrict predicted substitution patterns in helpful ways. For example, if a product removal coincides with idiosyncratically high overall levels of demand, a purely non-parametric analysis may imply that the removed product is a complement in consumption to other goods, even when this does not accord with basic economic intuition. Regardless of whether one uses a structural model, exogenous variation such as that induced by field experiments can help identify demand parameters by increasing the number of unique availability sets observed in the data. For example, Conlon and Mortimer estimate the structural demand models using (i) the full dataset, which contains 1 734 unique availability sets, and (ii) a subset of the data that excludes experimentally-induced variation, leaving only 1 096 unique availability sets. They find significant differences in the estimated parameters, especially for product categories in which a large fraction of availability sets are experimentally induced. This illustrates that when variation in observed availability sets is limited, identification relies more heavily on the functional-form assumptions of the structural model.

3.6 Behavioral Effect Studies

A number of studies have examined the potential for inventories to directly affect demand levels. For example, Ioannou, Mortimer, and Mortimer (2011) finds that video rental stores’ inventories of individual movie titles have a significant and positive effect on demand for movies with the highest costs of inventory (and thus, the most binding capacity constraints). The authors point to several possible mechanisms for this effect, including direct avoidance of stockouts and quality signaling. Similarly, Cachon, Gallino, and Olivares (2013) finds that when a car dealer increases inventory of an already in-

⁴¹ Using data from the same field experiment, Conlon and Mortimer (2015) also uses variation in product availability to examine the impact of a supply-side contractual arrangement, and Conlon and Mortimer (2014) develops a method for experimentally analyzing the impact of proposed mergers.

stock sub-model, demand for that model decreases. They suggest that this could be due to the consumers' concerns about a car's availability during continued search.⁴²

Responses to stockouts can also differ depending on the observability of the stockout. Pizzi and Scarpi (2013) finds that consumers are more likely to intend to visit a store in the future when the stockout disclosure is made before choosing a product and when the retailer claims responsibility for the stockout. Laboratory experiments have also shown that stockouts can lead to a decrease in the outside good share because they can increase consumer urgency or signal product quality (see Ge, Messinger, and Li (2009)), and single-option aversion, in which the market share of a sole remaining available product decreases in response to a stockout of a second product (see Mochon (2013)). These substitutions to and from the outside good share violate the restrictions imposed by many structural models, and could motivate models that separate the purchase incidence decision from product choice conditional on purchase (e.g., Campo, Gijsbrechts, and Nisol (2003), Breugelmans, Campo, and Gijsbrechts (2006), and Ching, Erdem, and Keane (2009)). Finally, Goldfarb (2006) presents evidence that habit-formation can be influenced by stockouts.

4 Limited Consumer Information

In many retail settings, consumers do not consider the entire set of available products when they make their decision. Limited memory and recall ability, limited computational ability, or costly information acquisition can all be used to rationalize limited product consideration. Since workhorse models of discrete choice assume that consumers consider all available products, limited consideration can lead to bias in estimated demand parameters. To distinguish this form of variation in “availability” from that of stockouts or assortment choices, and in line with common terminology, we refer to the resulting consumer-specific choice set as the *consideration set*.⁴³ Common methods for accounting for unobserved, non-deterministic consideration within discrete choice models can be grouped into roughly three approaches: (i) adjusted-utility, in which non-linear functions of some observables are included in the linear index of the choice probability function, (ii) two-stage models of choice that contain probabilities for both product consideration and choice conditional on a consideration set, and (iii) structural search models of the formation process, not just the existence, of consideration sets.⁴⁴

In a two stage (consideration, then choice) model, choice probabilities may be written as $p_j = \sum_C p_C p_{j|C}$, where p_j is the probability of choosing j , p_C is the probability of consideration set C , and $p_{j|C}$ is the probability of choosing j conditional on C . The adjusted-utility approach to consideration is often motivated by the idea of a separate consideration set formation stage, but then the quantity that would essentially be the probability of product consideration is inserted into the linear index function of the choice stage, resulting in a single stage model.⁴⁵ For example, Bronnenberg and Vanhonacker (1996)

⁴² Zettelmeyer, Morton, and Silva-Risso (2006), which also studies this effect in the market for cars, refers to it as a *scarcity rent*.

⁴³ Some scholars distinguish between awareness and consideration, where the former could mean that the consumer knows the product exists, but is unaware of certain products characteristics, such as price.

⁴⁴ Early reviews of the consideration set literature can be found in Roberts and Lattin (1997) and Manrai and Andrews (1998). Hauser (2014) summarizes work on inferring consumer consideration heuristics from *observed* consideration sets.

⁴⁵ We adapt the term “adjusted utility” from Wu and Rangaswamy (2003, p. 415), who describe their model as “search-adjusted choice utility.”

presents a model of the probability that a product's salience, s , exceeds a threshold θ required for consideration, given by $\pi_j = 1 / [1 + \exp(\theta + s_j)]$, where s_j is a linear function of observables. Note that this is the functional form of a binary logit model, so π takes values between one and zero. Bronnenberg and Vanhonacker then incorporate π_j into the choice model and write the multinomial probability of choosing product j as $p_j = [\pi_j \exp(\delta_j)] / [\sum_k \pi_k \exp(\delta_k)]$. Because the model consists of a single stage, it could also be interpreted as simply including a particular nonlinear function of certain observables in the linear index of the MNL form, as seen when one rewrites $\pi_j \exp(\delta_j)$ as $\exp(\ln \pi_j + \delta_j)$.⁴⁶

A larger body of literature explicitly models both a consideration and a product choice (conditional on consideration) stage. For many of these models, each product again has a particular *salience* or *consideration utility*, which is incorporated into a marginal probability of product consideration. The probability of a particular consideration set is then the joint probability of considering exactly those products contained in the set (e.g., Barroso and Llobet (2012), Pancras (2010), Goeree (2008), and Swait and Erdem (2007)). Most such models treat the joint probability of considering a set of products as the product of the marginal probabilities of consideration, i.e., they assume that consideration is conditionally independent across products.⁴⁷ Exceptions to the independence assumption include van Nierop et al. (2010), which assumes joint normality of the consideration errors and allow for a non-diagonal covariance matrix, and Dehmamy and Otter (2014), which writes out a full conditional (on all other products' consideration status) distribution for the consideration of product j .⁴⁸ Another approach to avoiding the independence assumption is to forgo the foundation of individual product salience and instead write multinomial choice probabilities for consideration sets as a function of their attributes (e.g., Draganska and Klapper (2011)).

Since the number of possible consideration sets increases exponentially in the number of products, direct estimation of the likelihood for more than a few products becomes computationally infeasible. A common solution to this problem is to use Markov Chain Monte Carlo methods to simulate consideration sets, as introduced in Chiang, Chib, and Narasimhan (1999) (see Goeree (2008) for another example). In addition to simulation, feasible computation may require numerical techniques to smooth the likelihood function so that it can be optimized. Honka (2014) uses kernel smoothing as an alternative to the prohibitively large number of draws that would be required to avoid lumpiness in the simulated probabilities.

As consideration sets are generally not observed, one could rationalize differences in product choice across choice sets as being due to either utility or consideration, or some combination. Many models of consideration sets rely heavily on assumed functional form (e.g., error distributions and exclusion

⁴⁶ Wu and Rangaswamy (2003) use a different non-linear function form for π_j , motivated by the concept of two fuzzy real numbers (e.g., salience and a threshold) and the fuzzy order of those numbers, instead of two real random variables and the probability that one is greater than another. Pancras (2011) adjusts utility for salience as in Bronnenberg and Vanhonacker (1996), but does so in both a store choice stage and a product choice stage.

⁴⁷ Barroso and Llobet (2012, p. 777) point out that this assumption "rules out the existence of cognitive constraints (e.g., a consumer can only remember a certain number of products) or information spillovers (e.g., a consumer aware of the existence of a product is also aware of the existence of other products sold by the same firm)."

⁴⁸ Given the large number of products in their application, calculation of the full joint density of each consideration set would be computationally infeasible. To overcome this challenge, they employ Gibbs sampling of the consideration set in their estimation.

restrictions) to separately identify the consideration parameters from the utility parameters. New identification approaches have appeared in more recent work. Dehmamy and Otter (2014) investigates the role that observed *quantity* choices can play in identifying consideration parameters separately from utility parameters. Kawaguchi, Uetake, and Watanabe (2014) describes how observed availability variation (e.g., stockouts), considered as exogenous variation in consideration sets, can enable identification of consideration parameters. In a further move away from reliance on parametric form for consideration, Lu (2014) develops a moment inequality estimator for demand with consideration, based on (i) *a priori* restrictions on the choice set (e.g., consumers consider all products on display and cannot consider unavailable products) and (ii) monotonicity of choice (i.e., if j is chosen from consideration set C , then j will be chosen from any subset of C that contains j). The advantages of this moment inequality approach are that it does not require specification of a fully parametric consideration model (thereby limiting misspecification error) and it reduces the computational burden because it doesn't require a calculation for each possible consideration set.⁴⁹

In addition to modeling the existence of consideration sets, some researchers have developed models of the process of consideration set formation using structural search. In structural models of simultaneous search consumers choose a set of products for which they will search for particular information (e.g., prices), based on search costs and expectations of the outcome of the search. The information obtained through search could be prices (e.g., Honka (2014), Seiler (2013), and Mehta, Rajiv, and Srinivasan (2003)), unobserved utility errors (e.g., Moraga-González, Sándor, and Wildenbeest (2012), or both (e.g., Spence (2014), Muir, Seim, and Vitorino (2013), and Pires (2012)). The chosen set of products constitutes the consideration set, from which the consumer then makes a choice after obtaining information.⁵⁰ One interesting feature of this research, as compared to the consideration literature without search underpinnings, is the selection of observed products on unobservables. For example, in Honka (2014) the consumer observes the random component of utility before deciding whether to incur costs to search for the price of the product. As a result of this selection, the distribution of the random component of utility (e.g., the i.i.d. TIEV error term of the MNL model) among considered products differs from the distribution in the set of all products. The choice probabilities must be corrected to account for this selection. Structural search methods of consideration set formation have been applied to a variety of data structures, including aggregate markets (e.g., Moraga-González, Sándor, and Wildenbeest (2012)), household surveys (e.g., Honka (2014)), household scanner data (e.g., Mehta, Rajiv, and Srinivasan (2003), Seiler (2013), and Pires (2012)), online search records (e.g., Koulayev (2014)), online product-rank data (e.g., Kim, Albuquerque, and Bronnenberg (2010)), and regulatory administrative data (e.g., Muir, Seim, and Vitorino (2013)).

5 Conclusion

Demand estimates are identified from variation in choice sets. When using highly aggregated data, such as national annual sales, the notion of 'choice set variation' usually refers to variation in the characteristics of products (including price) over time or across markets. However, as more granular data

⁴⁹ Fox (2007) develops a semiparametric estimator of discrete-choice demand that can accommodate random, unobserved consideration.

⁵⁰ In some models consumers choose whether or not to search in an entire category, which is a binary choice, rather than choosing from the subsets of the full availability set (see Seiler (2013) and Spence (2014)).

continue to become available in a wide range of markets, researchers are able to take advantage of new forms of ‘choice set variation’ that arise through variation in the *actual* choice sets faced by consumers. Sometimes this variation is endogenous, as in the case of product assortment. Some forms of this variation may plausibly arise exogenously, such as during stockout events. These sources of variation may arise across stores, markets, time, or even individual consumers. In each case, researchers must think carefully about how the choice set variation arises, its likely impact on estimated substitution patterns, and the implications for identification.⁵¹

Great progress has been made in recognizing the role that product availability plays in demand estimation, as well as incorporating various forms of availability into estimates of substitution patterns and demand elasticities for many types of data structures. Two important directions for future work include the development of additional methodological advances, and the application of existing methods to significant research questions in several fields.

Future methodological work could fruitfully combine product availability with other aspects of consumer choice. For example, the connection between availability and store choice is one for which further work is likely to be useful. Similarly, availability plays a key role in the demand for seasonal goods. Developing methods that accommodate availability in systems of differentiated seasonal goods would be quite helpful in these markets.

Methodological advances notwithstanding, the most important role of demand estimation is to provide crucial inputs (i.e., substitution patterns and elasticities) that enable researchers to address questions about optimal strategic behavior, and to assess the welfare effects of equilibrium responses to market environments and economic policies. Future work could apply existing methods to better understand, for example, how assortments, stockout frequencies, and consideration sets are shaped by vertical arrangements between retailers and manufacturers (e.g., slotting fees and promotional funds) in order to understand the impact of these arrangements for consumers. To this end, the potential for future research is extensive.

6 References

- Aguirregabiria, V. (2005), 'Retail stockouts and manufacturer brand competition', Working Paper.
- Aguirregabiria, V. (1999), 'The dynamics of markups and inventories in retailing firms', *The Review of Economic Studies*, **66** (2), 275-308.
- Andersen Consulting (1996), 'Where to look for incremental sales gains: The problem of out-of-stock merchandise': Coca-Cola Retailing Research Council, available at <http://www.ccrcc.org/1996/02/24/where-to-look-for-incremental-sales-gains-the-retail-problem-of-out-of-stock-merchandise/> (accessed 26 March 2015).
- 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.
- Azadeh, S. S., P. Marcotte, and G. Savard (2014), 'A taxonomy of demand uncensoring methods in revenue management', *Journal of Revenue and Pricing Management*, **13** (6), 440-456.
- Barnett, W. A., and A. Serletis (2008), 'Consumer preferences and demand systems', *Journal of Econometrics*, **147** (2), 210-224.
- Barroso, A., and G. Llobet (2012), 'Advertising and consumer awareness of new, differentiated products', *Journal of Marketing Research*, **49** (6), 773-792.
- Berry, S. T. (1994), 'Estimating discrete-choice models of product differentiation', *The RAND Journal of Economics*, **25** (2), 242-262.
- Berry, S. T., and P. Jia (2010), 'Tracing the woes: An empirical analysis of the airline industry', *American Economic Journal: Microeconomics*, **2** (3) 1-43.
- Berry, S. T., and J. Waldfogel (2001), 'Do mergers increase product variety? Evidence from radio broadcasting', *The Quarterly Journal of Economics*, **116** (3), 1009-1025.
- Berry, S. T., J. Levinsohn, and A. Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica*, **63** (4), 841-889.
- Besbes, O., and A. Muharremoglu (2013), 'On implications of demand censoring in the newsvendor problem', *Management Science*, **59** (6), 1407-1424.
- Boatwright, P., and J.C. Nunes (2001), 'Reducing assortment: An attribute-based approach', *Journal of Marketing*, **65** (3), 50-63.
- Borle, S., P. Boatwright, J. B. Kadane, J. C. Nunes, and G. Shmueli (2005), 'The effect of product assortment changes on customer retention', *Marketing Science*, **24** (4), 616-622.
- Breugelmans, E., K. Campo, and E. Gijsbrechts (2006), 'Opportunities for active stock-out management in online stores: The impact of the stock-out policy on online stock-out reactions', *Journal of Retailing*, **82** (3), 215-228.

- Briesch, R. A., W. R. Dillon, and R. C. Blattberg (2008), 'Treating zero brand sales observations in choice model estimation: Consequences and potential remedies', *Journal of Marketing Research*, **45** (5), 618-632.
- Broniarczyk, S. M., W. D. Hoyer, and L. McAlister (1998), 'Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction', *Journal of Marketing Research*, **35** (2), 166-176.
- Bronnenberg, B. J., and W. R. Vanhonacker (1996), 'Limited choice set, local price response and implied measures of price competition', *Journal of Marketing Research*, **33** (2), 163-173.
- Bronnenberg, B. J., S. K. Dhar, and J.P. H. Dubé (2007), 'Consumer packaged goods in the united states: National brands, local branding', *Journal of Marketing Research*, **44** (1), 4-13.
- Bronnenberg, B. J., V. Mahajan, and W. R. Vanhonacker (2000), 'The emergence of market structure in new repeat-purchase categories: The interplay of market share and retailer distribution', *Journal of Marketing Research*, **37** (1), 16-31.
- Bruno, H. A., and N. J. Vilcassim (2008), 'Structural demand estimation with varying product availability', *Marketing Science*, **27** (6), 1126-1131.
- Cachon, G. P., S. Gallino, and M. Olivares (2013), 'Does adding inventory increase sales? Evidence of a scarcity effect in U.S. automobile dealerships', available at <http://ssrn.com/abstract=2286800> (accessed 26 March 2015).
- Campo, K., E. Gijsbrechts, and P. Nisol (2004), 'Dynamics in consumer response to product unavailability: Do stock-out reactions signal response to permanent assortment reductions', *Journal of Business Research*, **57** (8), 834-843.
- Campo, K., E. Gijsbrechts, and P. Nisol (2003), 'The impact of retailer stockouts on whether, how much, and what to buy', *International Journal of Research in Marketing*, **20** (3), 273-286.
- Che, H., X. Chen, and Y. Chen (2012), 'Investigating effects of out-of-stock on consumer stockkeeping unit choice', *Journal of Marketing Research*, **49** (4), 502-513.
- Chen, L., and A. J. Mersereau (2014), 'Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy', available at https://faculty.fuqua.duke.edu/~lc91/More/papers/Chen_Mersereau_Analytics_2014.pdf (accessed 26 March 2015).
- Chiang, J., S. Chib, and C. Narasimhan (1999), 'Markov chain Monte Carlo and models of consideration set and parameter heterogeneity', *Journal of Econometrics*, **89** (1), 223-248.
- Ching, A., T. Erdem, and M. Keane (2009), 'The price consideration model of brand choice', *Journal of Applied Econometrics*, **24** (3), 393-420.

- Chintagunta, P., J.P. Dubé, and K. Y. Goh (2005), 'Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models', *Management Science*, **51** (5), 832-849.
- Conlon, C. T., and J. H. Mortimer (2014), 'An experimental approach to merger evaluation', Working Paper (see also NBER Working Paper 19703).
- Conlon, C. T., and J. H. Mortimer (2013), 'Demand estimation under incomplete product availability', *American Economics Journal: Microeconomics*, **5** (4), 1-30.
- Conlon, C. T., and J. H. Mortimer (2010), 'Effects of product availability: Experimental evidence', NBER Working Paper 16506.
- Conlon, C. T., and J. H. Mortimer (2015), 'Efficiency and foreclosure effects of all units discounts: Empirical evidence', Working Paper (see also NBER Working Paper 19709).
- Crawford, G. S. (2012), 'Endogenous product choice: A progress report', *International Journal of Industrial Organization*, **30** (3), 315-320.
- Dehmamy, K., and T. Otter (2014), 'Utility and attention-a structural model of consideration', available at <http://ssrn.com/abstract=2433145> (accessed 26 March 2015).
- DeHoratius, N., A. J. Mersereau, and L. Schrage (2008), 'Retail inventory management when records are inaccurate', *Manufacturing and Service Operations Management*, **10** (2), 257-277.
- 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, M., and D. C. Jain (2005), 'Product-line length as a competitive tool', *Journal of Economics and Management Strategy*, **14** (1), 1-28.
- Draganska, M., and D. Klapper (2011), 'Choice set heterogeneity and the role of advertising: An analysis with micro and macro data', *Journal of Marketing Research*, **48** (4), 653-669.
- Draganska, M., M. Mazzeo, and K. Seim (2009), 'Beyond plain vanilla: Modeling joint product assortment and pricing decisions', *Quantitative Marketing and Economics*, **7** (2), 105-146.
- Eizenberg, A. (2014), 'Upstream innovation and product variety in the US home PC market', *The Review of Economic Studies*, **81** (3), 1003-1045.
- Escobari, D. (2012), 'Dynamic pricing, advance sales and aggregate demand learning in airlines', *The Journal of Industrial Economics*, **60** (4), 697-724.
- Figurelli, L. (2013), 'Store choice in spatially differentiated markets', available at https://www2.bc.edu/lucrezio-figurelli/Site/Home_files/jmp_figurelli.pdf (accessed 26 March 2015).
- Fox, J. T. (2007), 'Semiparametric estimation of multinomial discrete-choice models using a subset of choices', *The RAND Journal of Economics*, **38** (4), 1002-1019.

- Gandhi, A., Z. Lu, and X. Shi (2014), 'Demand estimation with scanner data: Revisiting the loss-leader hypothesis', available at http://economics.uchicago.edu/workshops/pdfs/Gandhi_Scanner_Demand_v2.pdf (accessed 26 March 2015).
- Gandhi, A., Z. Lu, and X. Shi (2013), 'Estimating demand for differentiated products with error in market shares', CEMMAP Working Paper CWP03/13, London: Centre for Microdata Methods and Practice.
- Ge, X., P. R. Messinger, and J. Li (2009), 'Influence of soldout products on consumer choice', *Journal of Retailing*, **85** (3), 274-287.
- Goeree, M. (2008), 'Limited information and advertising in the U.S. personal computer industry', *Econometrica*, **76** (5), 1017-1074.
- Goettler, R. L., and B. R. Gordon (2011), 'Does AMD spur Intel to innovate more? *Journal of Political Economy*, **119** (6), 1141-1200.
- Goldfarb, A. (2006), 'The medium-term effects of unavailability', *Quantitative Marketing and Economics*, **4** (2), 143-171.
- Gourville, J. T., and D. Soman (2005), 'Overchoice and assortment type: When and why variety backfires', *Marketing Science*, **24** (3), 382-395.
- Gowrisankaran, G., and M. Rysman (2012), 'Dynamics of consumer demand for new durable goods', *Journal of Political Economy*, **120** (6), 1173-1219.
- Gruen, T. W., D. Corsten, and S. Bharadwaj (2002), 'Retail out of stocks: A worldwide examination of causes, rates, and consumer responses' : Grocery Manufacturers of America, available at http://itsoutofstock.com/wp-content/uploads/2013/04/GMA_2002_-Worldwide_OOS_Study.pdf (accessed 26 March 2015).
- Guo, P., B. Xiao, and J. Li (2012), 'Unconstraining methods in revenue management systems: Research overview and prospects', *Advances in Operations Research* **2012**.
- Hauser, J. R. (2014), 'Consideration-set heuristics', *Journal of Business Research*, **67** (8), 1688-1699.
- Honka, E. (2014), 'Quantifying search and switching costs in the US auto insurance industry', *The RAND Journal of Economics*, **45** (4), 847-884.
- Hwang, M., and R. Thomadsen (Forthcoming), 'How point-of-sale marketing mix impacts national-brand purchase shares', *Management Science*.
- Hwang, M., B. J. Bronnenberg, and R. Thomadsen (2010), 'An empirical analysis of assortment similarities across U.S. supermarkets', *Marketing Science*, **29** (5), 858-879.
- Iaria, A. (2014), 'Consumer benefit of big-box supermarkets: The importance of controlling for endogenous entry', Mimeo.
- Ioannou, I., J. H. Mortimer, and R. Mortimer (2011), 'The effects of capacity on sales under alternative vertical contracts', *Journal of Industrial Economics*, **59** (1), 117-154.

- Jiang, R., P. Manchanda, and P. E. Rossi (2009), 'Bayesian analysis of random coefficient logit models using aggregate data', *Journal of Econometrics*, **149** (2), 136-148.
- Karabati, S., B. Tan, and Ö. C. Öztürk (2009), 'A method for estimating stock-out-based substitution rates by using point-of-sale data', *IIE Transactions*, **41** (5), 408-420.
- Katz, M. (2007), 'Estimating supermarket choice using moment inequalities' : Harvard University, Mimeo.
- Kawaguchi, K., K. Uetake, and Y. Watanabe (2014), 'Identifying consumer inattention: A product-availability approach', available at <http://ssrn.com/abstract=2529294> (accessed 26 March 2015).
- Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010), 'Online demand under limited consumer search', *Marketing science*, **29** (6), 1001-1023.
- Kök, A. G., and M. L. Fisher (2007), 'Demand estimation and assortment optimization under substitution: Methodology and application', *Operations Research*, **55** (6), 1001-1021.
- Kök, A. G., M. L. Fisher, and R. Vaidyanathan (2009), 'Assortment planning: Review of literature and industry practice', In N. Agrawal and S. Smith (Eds.), *Retail Supply Chain Management: Quantitative Models and Empirical Studies*, New York: Springer, 99-153.
- Koulayev, S. (2014), 'Search for differentiated products: Identification and estimation', *The RAND Journal of Economics*, **45** (3), 553-575.
- Kramer, T., and R. Carroll (2009), 'The effect of incidental out-of-stock options on preferences', *Marketing Letters*, **20** (2), 197-208.
- Lazarev, J. (2013), 'The welfare effect of intertemporal price discrimination: An empirical analysis of airline pricing in US monopoly markets', available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.5942&rep=rep1&type=pdf> (accessed 26 March 2015).
- Lee, H., and Ö. Özer (2007), 'Unlocking the value of RFID', *Production and Operations Management*, **16** (1), 40-64.
- Leslie, P. (2004), 'Price discrimination in Broadway theater', *The RAND Journal of Economics*, **35** (3), 520-541.
- Leslie, P., and A. Sorensen (2014), 'Resale and rent-seeking: An application to ticket markets', *Review of Economic Studies*, **81** (1), 266-300.
- Li, B., A. E. Lim, and T. Wang (2014), 'Estimation and optimization of logit demand model with covariates, missing data, and auxiliary information', available at <http://bschool.nus.edu/staff/bizwt/LiLimWang.pdf> (accessed 26 March 2015).
- Lu, Z. (2014), 'A moment inequality approach to estimating multinomial choice models with unobserved consideration sets', available at http://www.ssc.wisc.edu/~zlu33/job_market/jmp.pdf (accessed 26 March 2015).

- Manchanda, P., P. E. Rossi, and P. K. Chintagunta (2004), 'Response modeling with nonrandom marketing-mix variables', *Journal of Marketing Research*, **41** (4), 467-478.
- Manrai, A. K., and R. L. Andrews (1998), 'Two-stage discrete choice models for scanner panel data: An assessment of process and assumptions', *European Journal of Operational Research*, **111** (2), 193-215.
- Mantrala, M. K., M. Levy, B. E. Kahn, E. J. Fox, P. Gaidarev, B. Dankworth, et al. (2009), 'Why is assortment planning so difficult for retailers? A framework and research agenda', *Journal of Retailing*, **85** (1), 71-83.
- Matsa, D. A. (2011), 'Competition and product quality in the supermarket industry', *The Quarterly Journal of Economics*, **126** (3), 1539-1591.
- Mehta, N., S. Rajiv, and K. Srinivasan (2003), 'Price uncertainty and consumers search: A structural model of consideration set formation', *Marketing Science*, **22** (1), 58-84.
- Misra, K. (2008), 'Understanding retail assortments in competitive markets', available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.181.288&rep=rep1&type=pdf> (accessed 26 March 2015).
- Mochon, D. (2013), 'Single-option aversion', *Journal of Consumer Research*, **40** (3), 555-566.
- Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2011), 'Consumer search and prices in the automobile market', available at <http://www.unc.edu/~joubertc/searchautomobilesUNC.pdf> (accessed 26 March 2015).
- Muir, D., K. Seim, and M. Vitorino (2013), 'Price obfuscation and consumer search: An empirical analysis', available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.1338&rep=rep1&type=pdf> (accessed 26 March 2015).
- Musalem, A. (2015), 'When demand projections are too optimistic: A structural model of product line and pricing decisions', available at <http://ssrn.com/abstract=2560662> (accessed 26 March 2015).
- Musalem, A., M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten (2010), 'Structural estimation of the effect of out-of-stocks', *Management Science*, **56** (7), 1180-1197.
- Nevo, A. (2000), 'A practitioner's guide to estimation of random-coefficients logit models of demand', *Journal of Economics and Management Strategy*, **9** (4), 513-548.
- Nevo, A. (2011), 'Empirical models of consumer behavior', *Annual Review of Economics*, **3** (1), 51-75.
- Newman, J. P., M. E. Ferguson, L. A. Garrow, and T. L. Jacobs (2014), 'Estimation of choice-based models using sales data from a single firm', *Manufacturing and Service Operations Management*, **16** (2), 184-197.
- Pakes, A. (2003), 'A reconsideration of hedonic price indexes with an application to PC's', *American Economic Review*, **93** (5), 1578-1596.

- Pakes, A., and T. Erickson (2011), 'An experimental component index for the CPI: From annual computer data to monthly data on other goods', *American Economic Review*, **101** (5), 1707-38.
- Pancras, J. (2010), 'A framework to determine the value of consumer consideration set information for firm pricing strategies', *Computation Economics*, **35** (3), 269-300.
- Pancras, J. (2011), 'The nested consideration model: Investigation dynamic store consideration sets and store competition', *European Journal of Operational Research*, **214** (2), 340-347.
- Pentico, D. W. (2008), 'The assortment problem: A survey', *European Journal of Operational Research*, **190** (2), 295-309.
- Petrin, A., and K. Train (2010), 'A control function approach to endogeneity in consumer choice models', *Journal of Marketing Research*, **47** (1), 3-13.
- Pires, T. (2012), 'Consideration Sets in Storable Goods Markets', available at <http://gradstudents.wcas.northwestern.edu/~tmp767/ConsiderationSets.pdf> (accessed 26 March 2015).
- Pizzi, G., and D. Scarpi (2013), 'When out-of-stock products DO backfire: Managing disclosure time and justification wording', *Journal of Retailing*, **89** (3), 352-359.
- Quan, T. W., and K. R. Williams (2014), 'Product variety, across-market demand heterogeneity, and the value of online retail', available at <http://www.econ.umn.edu/~quanx039/quantJMP.pdf> (accessed 26 March 2015).
- Raman, A., N. DeHoratius, and Z. Ton (2001), 'Execution: The missing link in retail operations', *California Management Review*, **43** (3), 136-151.
- Reibstein, D. J., and P. W. Farris (1995), 'Market share and distribution: A generalization, a speculation, and some implications', *Marketing Science*, **14** (3), G190-G202.
- Rhee, H., and D. R. Bell (2002), 'The inter-store mobility of supermarket shoppers', *Journal of Retailing*, **78** (4), 225-237.
- Richards, T. J., and S. F. Hamilton (2006), 'Rivalry in price and variety among supermarket retailers', *American Journal of Agricultural Economics*, **88** (3), 710-726.
- Roberts, J. H., and J. M. Lattin (1997), 'Consideration: Review of research and prospects for future insights', *Journal of Marketing Research*, **34** (3), 406-410.
- Seiler, S. (2013), 'The impact of search costs on consumer behavior: A dynamic approach', *Quantitative Marketing and Economics*, **11** (2), 155-203.
- Shah, D., V. Kumar, and Y. Zhao (Forthcoming), 'Diagnosing brand performance: Accounting for the dynamic impact of product availability with aggregate data', *Journal of Marketing Research*.
- Sloot, L. M. and P. C. Verhoef (2008), 'The impact of brand delisting on store switching and brand switching intentions', *Journal of Retailing*, **84** (3), 281-296.

- Sloot, L. M., D. Fok, and P. C. Verhoef (2006), 'The short- and long-term impact of an assortment reduction on category sales', *Journal of Marketing Research*, **43** (4), 536-548.
- Smith, S. A. and N. Agrawal (2000), 'Management of multi-item retail inventory systems with demand substitution', *Operations Research*, **48** (1), 50-64.
- Soysal, G. P. and L. Krishnamurthi (2012), 'Demand dynamics in the seasonal goods industry: An empirical analysis', *Marketing Science*, **31** (2), 293-316.
- Spence, F. (2014), 'Consumer experience and the value of search in the online textbook market', available at <http://forrestspence.web.unc.edu/files/2014/11/SpenceJMP.pdf> (accessed 26 March 2015).
- Stavrulaki, E. (2011), 'Inventory decisions for substitutable products with stock-dependent demand', *International Journal of Production Economics*, **129** (1), 65-78.
- Swait, J. and T. Erdem (2007), 'Brand effects on choice and choice set formation under uncertainty', *Marketing Science*, **26** (5), 679-697.
- Sweeting, A. (2012), 'Dynamic pricing behavior in perishable goods markets: Evidence from secondary markets for major league baseball tickets', *Journal of Political Economy*, **120** (6), 1133-1172.
- Talluri, K. and G. van Ryzin (2004), 'Revenue management under a general discrete choice model of consumer behavior', *Management Science*, **50** (1), 15-33.
- Tan, B. and S. Karabati (2013), 'Retail inventory management with stock-out based dynamic demand substitution', *International Journal of Production Economics*, **145** (1), 78-87.
- Tenn, S. (2009), 'Demand estimation under limited product availability', *Applied Economics Letters*, **16** (5), 465-468.
- Tenn, S. and J. M. Yun (2008), 'Biases in demand analysis due to variation in retail distribution', *International Journal of Industrial Organization*, **26**, 984-997.
- Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge, UK and New York, NY, USA: Cambridge University Press.
- van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010), 'Retrieving unobserved consideration sets from household panel data', *Journal of Marketing Research*, **47**, 63-74.
- van Ryzin, G. and G. Vulcano (2014), 'A market discovery algorithm to estimate a general class of nonparametric choice models', *Management Science* **61** (2), 281-300.
- Vulcano, G., G. van Ryzin, and W. Chahr (2010), 'OM practice - Choice-based revenue management: An empirical study of estimation and optimization', *Manufacturing and Service Operations Management*, **12** (3), 371-392.
- Vulcano, G., G. van Ryzin, and R. Ratliff (2012), 'Estimating primary demand for substitutable products from sales transaction data', *Operations Research*, **60** (2), 313-334.

Williams, K. R. (2013), 'Dynamic airline pricing and seat availability', available at <http://kelley.iu.edu/BEPP/documents/williams%20paper.pdf> (accessed 26 March 2015).

Wu, J. and A. Rangaswamy (2003), 'A fuzzy set model of search and consideration with an application to an online market', *Marketing Science*, **22** (3), 411-434.

Yang, S., Y. Chen, and G. M. Allenby (2003), 'Bayesian analysis of simultaneous demand and supply', *Quantitative Marketing and Economics*, **1** (3), 251-275.

Zettelmeyer, F., F. S. Morton, and J. Silva-Risso (2006), 'Scarcity rents in car retailing: Evidence from inventory fluctuations at dealerships', NBER Working Paper 12177.