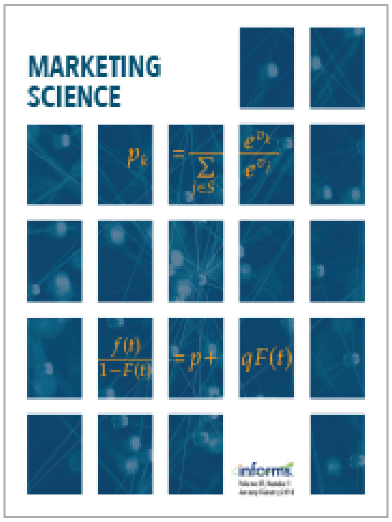


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Research Note—Structural Demand Estimation with Varying Product Availability

Hernán A. Bruno, Naufel J. Vilcassim,

MARKETING SCIENCE
Vol. 27, No. 6, November–December 2008, pp. 1126–1131
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2706 | 1126

DOI 10.1287/mksc.1080.0366

Research Note

Structural Demand Estimation with Varying Product Availability

Hernán A. Bruno
Erasmus School of Economics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands, bruno@few.eur.nl

Naufel J. Vilcassim
London Business School, London NW1 4SA, United Kingdom, nvilcassim@london.edu

This paper develops a model that extends the traditional aggregate discrete-choice-based demand model (e.g. Berry et al. 1995) to account for varying levels of product availability. In cases where not all products are available at every consumer shopping trip, the observed market share is a convolution of two factors: consumer preferences and the availability of the product in stores. Failing to account for the varying degree of availability would produce incorrect estimates of the demand parameters. The proposed model uses information on aggregate availability to simulate the potential assortments that consumers may face in a given shopping trip. The model parameters are estimated by simulating potential product assortment vectors by drawing multivariate Bernoulli vectors consistent with the observed aggregate level of availability. The model is applied to the UK chocolate confectionery market, focusing on the convenience store channel. We compare the parameter estimates to those obtained from not accounting for varying availability and analyze some of the substantive implications.

Key words: structural modeling; availability; retailing
History: This paper was received April 28, 2006, and was with the authors 11 months for 2 revisions; processed by Tulin Erdem. Published online in *Articles in Advance* June 19, 2008.

1. Introduction

The development of demand estimation methods based on individual-level discrete-choice models using aggregate data has allowed researchers to infer the distribution of consumer preferences by observing aggregate realizations of demand, product and demographic characteristics, and other external factors affecting demand (Berry et al. 1995). One key underlying assumption of aggregate choice models is the complete availability of all products over the consumer choice occasions. Although this assumption may be realistic in certain choice situations (e.g., mobile telephone services), it may not hold in others such as grocery products. With market-level data (cf. Dube 2004), where sales are aggregated across households and stores, part of the observed variation in market shares could be explained by variation in distribution coverage, as well as assortment variation or stockouts in different retailers or zones. In such cases, one would be ascribing differences in market shares to variations in consumers' taste and behaviors, even though there might be a significant contribution to the observed variance from different levels of availability.[1]

One of the reasons aggregate availability varies is that retailers make assortment decisions. Empirical studies have shown a convex relationship between distribution and market share (Reibstein and Farris 1995). Additionally, availability can vary because of out of stocks or slow adoption of new products (Bronnenberg et al. 2000, Bronnenberg and Mela 2004). Whatever the reason(s) for the variation in availability, two important questions are: What impact does it have on estimated demand effects of marketing mix and other variables? Would ignoring it lead to incorrect inferences about demand effects?

The issue of varying product availability has been absent from the reviews of the literature on structural

[1] Demand estimation with varying levels of availability has been proposed in the past. For example, Jeuland (1979) integrated a model of availability into a model of multibrand choices. In discrete-choice random utility methods, varying availability is corrected by introducing the corresponding *observed* choice set (e.g., Campo et al. 2003).

modeling (see Dube et al. 2005, Chintagunta et al. 2006, Kadiyali et al. 2001). One possible exception to this is the work of Tenn (2006). Although Tenn's (2006) work deals with store heterogeneity in promotions, its model could be applied to the problem of varying product availability provided some of the assumptions made there are applicable. Specifically, Tenn's (2006) analysis uses the fact that most of the time only one product is promoted in the product category analyzed (premium ice cream), which allows him to identify the marginal distribution of promotional activity. Applying this assumption to varying product availability would require that at any point in time only a single product has less than 100% availability—something that clearly does not hold in our data. An additional assumption is that the promotions are the same across all stock-keeping units (SKUs) of a given brand. In many categories where size or product form effects are present (for example, when only one size (32 oz.) or one form (powder) of a detergent is on promotion), so that aggregating across SKUs to the brand level is not appropriate, applying Tenn's (2006) model would be computationally challenging, if not infeasible. We contend that the method proposed here would be more suitable for the situation with many products (both brands and SKUs) in the category as it is computationally simpler. This paper presents a tractable model that can account for varying levels of product availability, and which can be easily estimated similar to Berry et al. (1995).

This paper is designed to (1) recognize that ignoring varying levels of availability when using aggregate-level data, as mentioned above, can lead to incorrect inferences about consumer preferences and, more important, (2) provide a method to use aggregate measures of product availability to statistically correct the demand estimates. Our model can be considered an extension of the well known discrete choice-based models with aggregate data. The main challenge is that, although we observe the aggregate level of distribution, we do not observe the set of products available in a particular store. Hence we use simulation methods to model the different possible assortments of products that individual stores may have had on a given shopping trip.

To illustrate the availability issue and apply our model, we analyze data from the chocolate confectionery industry sold through the impulse channel in the United Kingdom.

2. Accounting for Varying Levels of Availability

2.1. Observed and Conditional Market Shares

We model a market of J partially differentiated competing products. Consumers can choose any one of the J alternatives with respective prices $p_1, \ldots, p_J$ or choose not to purchase any (i.e., they have an outside option). At time t, the choice of alternative $j = 1, \ldots, J$ with a characteristics vector x_{jt} provides consumer h with a level of utility equal to $u_{hjt} = x_{jth}\beta_h - \alpha_h p_{jt} + \xi_{jt} + \varepsilon_{hjt}$.

The parameter vector β_h and the parameter α_h capture, respectively, the taste for the product characteristics and the price sensitivity for consumer h. The term ξ_{jt} captures unobserved product characteristics or unobserved market-level demand shocks, depending on the definition of x. We assume that the nopurchase option ($j = 0$) provides a utility level of zero. If ε is extreme value distributed, the probability choice follows the logit form (McFadden 1974).

Heterogeneity in preferences among consumers is modeled by assuming that the variation in individual parameters follows a probability distribution such as a continuous parametric distribution, or a nonparametric mass-point distribution. To simplify the notation, we write the individual-level parameters in terms of their mean across individuals and an individual-level deviation (noted as realization of the random variables ν_α and ν_β). Hence we write the parameters as $\alpha_h = \alpha + \sigma_\alpha \nu_\alpha$ and $\beta_h = \beta + \Sigma\nu_\beta$, where the scalar σ_α and the matrix Σ are parameters. Additionally, we write the mean utility level (that is, constant across consumers) as $\delta_{jt} = x_{jth}\beta_h - \alpha_h p_{jt} + \xi_{jt}$, and collect the heterogeneity terms in $\mu_{hjt} = x_{jt}\Sigma\nu_\beta - \sigma_\alpha\nu_\alpha p_{jt}$. Conditional on the parameters, the logit market share for product j is the integral over the heterogeneity $\nu = [\nu_\alpha, \nu_\beta]$,

$$\tilde{S}_{jt} = \int \frac{\exp[\delta_{jt} + \mu_{jt}(\nu)]}{1 + \sum_{k=1}^{J} \exp[\delta_{kt} + \mu_{kt}(\nu)]}\, dG(\nu; \sigma_\alpha, \Sigma). \quad (1)$$

The previous expressions, though ubiquitous in the literature, rely on the assumption that all products are available when consumers make purchase decisions. The notation $\tilde{S}_j$ represents the share *conditional on complete availability*, in contrast to the unconditional share S_j, defined below.

We let a_{jt} denote the availability of product j during a purchase occasion at time t. The variable a_{jt} can only take two values: 1 if the product is available and 0 if the product is not available. The set of products available is therefore denoted by the J-dimensional vector $\mathbf{a}_t = (a_{1t}, \ldots, a_{Jt})$. The expectation over preferences $\tilde{S}_{jt}(\mathbf{a}_t)$ is the share conditional on $\mathbf{a}_t$, and hence a function of the availability vector,

$$\tilde{S}_{jt}(\mathbf{a}_t) = \int \frac{a_{jt}\exp[\delta_{jt} + \mu_{jt}(\nu)]}{1 + \sum_{k=1}^{J} a_{kt}\exp[\delta_{kt} + \mu_{kt}(\nu)]}\, dG(\nu; \sigma_\alpha, \Sigma).$$

The observed market share is the expectation of the expression above over all possible availability vectors.

That is, the share can be written as an expectation over all values of a,

$$S_{jt} = \sum_{\text{all } a_t} \tilde{S}_{jt}(\mathbf{a}_t)\pi(\mathbf{a}_t). \tag{2}$$

In Equation (2) all $\mathbf{a}_t$ denotes all possible combinations of $a_{jt} \in \{0, 1\}$, and $\pi(\mathbf{a}_t)$ is the joint probability of observing the vector $\mathbf{a}_t$. In summary, the observed share is the result of averaging over all the shopping trips consumers made in period t, each of which involved the choice over a particular realization of the availability vector $\mathbf{a}_t$.

2.2. Estimation Strategy

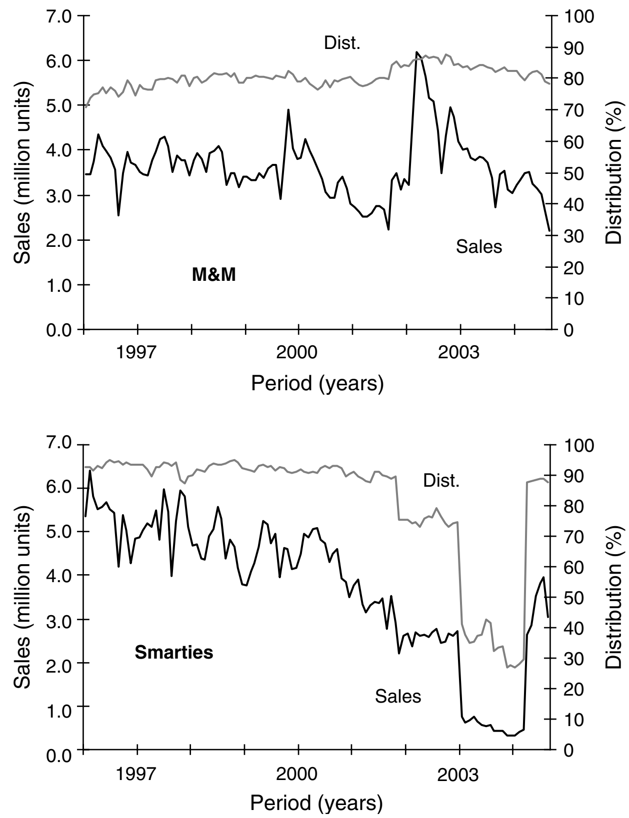
The proposed method introduces information on the observed levels of aggregate availability in the computation of the market shares. In essence, we use the aggregate level of availability to approximate the probability of finding a particular product, and thus compute the expectation in Equation (2). The logit estimation of demand involves finding parameters that minimize a convenient measure of distance between the observed and the computed market shares. Berry et al. (1995) proposed using an empirical distribution of pseudorandom draws to approximate the population density. The expectation over consumer preferences can be computed by summing over the random draws. Traditionally, each draw is a particular realization of the multivariate distribution of preferences. The method we propose involves generating vectors of product availability by taking random draws from the empirical distribution of availability. These availability vectors are then used in addition to the utility random draws to compute the market shares using Equation (2).

Note that the proposed approach does not involve putting additional parameters into the model. Our data include information that can be used as a proxy for the probability of availability for each product $\pi(a_1), \ldots, \pi(a_J)$ for each market time period. However, we do not observe the joint probability of the multivariate Bernoulli vector of availability $\pi(\mathbf{a}) = \pi(a_1, \ldots, a_J)$. To compute the joint probability, we make further assumptions about the correlation among the a_js. We assume that $\pi(a_j)$ is independent of $\pi(a_k)$ for $j = k$. That is to say, the probability of one product being available is independent of whether other products are available. This is an acceptable assumption in some situations and particularly in the context of our estimation below.

Let $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_J)$ denote the ith multivariate Bernoulli draw with marginal distributions given by $\pi(a_1), \ldots, \pi(a_J)$. We use this draw to approximate the sum of integrals in Equation (2). In the spirit of Berry et al. (1995) and Nevo (2001), we compute the sum

$$\hat{S}_{jt}(\delta) = \frac{1}{I}\sum_{i=1}^{I} \frac{\hat{a}_{jti}\exp[\delta_{jt} + \mu_{jt}(\hat{\nu}_i)]}{1 + \sum_{k=1}^{J} \hat{a}_{kti}\exp[\delta_{kt} + \mu_{kt}(\hat{\nu}_i)]}. \tag{3}$$

Figure 1 Sales and Distribution of Similar Products



By equating the share computed in Equation (3) to the observed shares, we obtained a system of equations analogous to those in Berry (1994), Berry et al. (1995), and Nevo (2001). It can be shown that the mean utility levels can be obtained using a recursive form as outlined in Berry (1994) and Berry et al. (1995).[2]

3. Application

We apply our estimation framework to characterize the demand system in the chocolate confectionery industry in the United Kingdom. We focus on the impulse channel (convenience stores, newsagents, tobacconists, etc.) where our assumptions are likely to hold.

Figure 1 illustrates the effect of distribution on sales. The drop in sales of Smarties seen around 2003 coincides with falling distribution coverage and should be noted if we are estimating consumer demand. Otherwise, the drop in demand will affect the demand parameter estimates and lead to incorrect inferences about consumer preferences and price sensitivities.

The data were provided by Information Resources, Inc. (IRI) and cover 113 four-week periods from June 1996 to February 2005 on confectionery products from

[2] The proof is available from the authors.

Table 1 **Summary of the Data Set**

Product	Manufacturer	Variable format	Weeks on market	Average availability (%)	Average price (GBP)	Average share (%)
Dairy Milk	Cadbury	Block	113	95	0.53	9.00
Whole Nut	Cadbury	Block	113	86	0.55	3.50
Fruit & Nut	Cadbury	Block	113	90	0.54	4.10
Caramel	Cadbury	Block	113	86	0.36	4.60
Flake	Cadbury	Indulgence	113	85	0.34	3.70
Snowflake	Cadbury	Indulgence	58	53	0.37	1.50
Dream	Cadbury	Block	109	29	0.66	0.70
Wispa	Cadbury	Indulgence	113	65	0.33	2.10
Galaxy	Masterfoods	Block	113	87	0.47	5.10
Galaxy Caramel	Masterfoods	Block	113	66	0.34	2.10
Galaxy Ripple	Masterfoods	Indulgence	113	70	0.32	2.70
M&Ms	Masterfoods	Bite sized	113	80	0.47	3.60
Maltesers	Masterfoods	Bite sized	113	95	0.54	8.00
Mars	Masterfoods	Filler	113	99	0.34	15.70
Mars 5 Little Ones	Masterfoods	Indulgence	53	14	0.35	0.30
Kit Kat	Nestle	Indulgence	113	96	0.34	11.80
Kit Kat Chunky	Nestle	Filler	76	89	0.35	6.70
Kit Kat Cubes	Nestle	Bite sized	18	63	0.50	1.60
Milky Bar	Nestle	Block	63	69	0.25	4.00
Milky Bar Munchies	Nestle	Bite sized	38	40	0.46	0.70
Smarties	Nestle	Bite sized	113	81	0.41	3.70
Double Cream	Nestle	Block	33	65	0.49	1.40
Snicker	Masterfoods	Filler	113	97	0.35	10.80
Snickers Cruncher	Masterfoods	Filler	45	38	0.32	0.70

the UK market. For each product, we observe national unit sales, average price, and category-weighted distribution. We defined the total market as the total units of confectionery consumed in the corresponding time period. We constructed the market share as the ratio of sales to the total market unit sales. In measuring aggregate availability, we use the all commodity volume (ACV) weighted distribution as the aggregate availability variable to recognize that not all stores have identical sales and traffic. Although this is not the ideal measure of distribution (Dolan and Hayes 2005), it serves as a good proxy considering that we are using national UK-level data. In addition, ACV is a standard measure used by practitioners and researchers and is widely available from market research companies. Note also that the data (including distribution) are aggregated at the brand level (see Little 1998 for a discussion of aggregation for sales and distribution measures). This is a common procedure when the estimation is performed at the product level (as opposed to using a characteristics-based approach). We expect these shortcomings (common to many situations in which aggregate data are used) to be minimized by focusing on the impulse channel because these smaller stores are less likely to carry a large variety of SKUs.[3]

We use four variables that indicate the format of the confectionery product (block, filler, indulgence, and bite sized). In Table 1, we provide a summary of the data set and we can see that most of the products were available for the whole period of observation. Products that were not available during the whole period of observation were launched after June 1996. No product in the data set disappeared during the period of observation. The confectionery market is composed of thousands of different products with many different variants and SKUs. The data set we are using aggregates across different SKUs and, overall, covers 70% of the market. Because of the highly fragmented nature of the category, small market shares (on the order of 1%) do not necessarily represent low levels of sales. Finally, note that some products have significant market shares in spite of low availability levels (e.g., Snowflake). It should be noted from Table 1 that some products have high levels of availability, although they are not in the market in the early weeks because they were launched after June 1996. Conversely, other products are present in the market for the whole period observation but show low levels of availability.

To account for potential endogeneity, we use three types of instruments for price, as it is well accepted

[3] If they stock a product, they tend to carry the single-pack, regular size, regular flavor SKU. We found that this is the case by observing data at the SKU level for the years 2002–2005. In addition, stores in the impulse channels are similar to one another in terms of sales volume and size, thus minimizing the aggregation bias in the ACV computation.

Table 2 Product Intercepts for the Standard and the Proposed Method

Product	Estimates (standard errors) Random coefficients logit	Random coefficients logit with/availability
Dairy Milk	−1.047 (0.606)	−1.130 (1.021)
Whole Nut	−1.874 (0.604)	−1.932 (0.177)
Fruit and Nut	−1.726 (0.595)	−1.794 (0.264)
Caramel	−2.488 (0.413)	−2.477 (0.189)
Flake	−2.863 (0.383)	−3.140 (0.375)
Snowflake	−3.626 (0.657)	−3.664 (1.021)
Dream	−4.236 (1.210)	−3.377 (0.177)
Wispa	−3.224 (0.471)	−3.455 (0.264)
Galaxy	−1.848 (0.533)	−1.853 (0.189)
Galaxy Caramel	−3.485 (0.444)	−3.194 (0.375)
Galaxy Ripple	−3.193 (0.360)	−3.273 (1.021)
M&Ms	−4.716 (0.586)	−4.994 (0.177)
Maltesers	−3.572 (0.632)	−3.974 (0.264)
Mars	−1.632 (0.396)	−2.212 (0.189)
Mars 5 Little Ones	−9.403 (4.587)	−7.924 (0.375)
Kit Kat	−1.565 (0.426)	−1.843 (1.021)
Kit Kat Chunky	−2.507 (0.554)	−3.054 (0.177)
Kit Kat Cubes	−5.322 (1.095)	−6.018 (0.264)
Milky Bar	−2.780 (0.634)	−2.828 (0.189)
Milky Bar Munchies	−6.581 (0.892)	−6.486 (0.375)
Smarties	−5.022 (0.516)	−5.479 (1.021)
Double Cream	−3.097 (0.731)	−3.000 (0.177)
Snicker	−1.987 (0.413)	−2.579 (0.264)
Snickers Cruncher	−5.185 (1.548)	−5.353 (0.189)

that the unobserved term ξ_{jt} is likely to be correlated with the price of the product. First, we use price indices for labor costs and spot prices for cocoa. Second, we use average prices from our manufacturer products (as they may capture manufacturer-specific changes in marginal cost). Finally, we use average prices of products of the same format. These variables were selected after carefully studying their explanatory power on price using hedonic regressions. Together, they explain around 80% of the observed variance in prices.

It can be seen from the results in Table 2, that when availability is corrected for using the proposed methodology, two-thirds of the product intercept estimates are higher. As we pointed out in the introduction, not accounting for availability penalizes products with low distribution coverage because of, for instance, sluggish retailer adoption. In contrast, well-established products with very high distribution and market shares (Mars, Snickers, Kit Kat, and popular plain chocolates) have lower product intercept estimates.

In Table 3, we show the estimates for the product characteristics.

These patterns also translate to the product characteristics, although the format coefficients are harder to interpret because they depend on many products. Note, however, that controlling for availability produces different estimates, particularly in the price coefficient, which is smaller in absolute value (i.e., less price sensitivity). In many cases, the best-selling brands in a given category have high levels of availability. If the estimation were restricted to the top few brands in the category, the results from the two methods would not differ much from each other. However, in the category we are analyzing, top brands account for only a fraction of total category sales. So estimating the model using only the top brands (a practice that could be sensible when analyzing other categories such as ketchup) would leave a large proportion of the market out of the estimation. In general, we would expect this to be the case for situations in which a large number of products have to be included in the estimation (e.g., the Nevo 2001 analysis of the cereal product category).

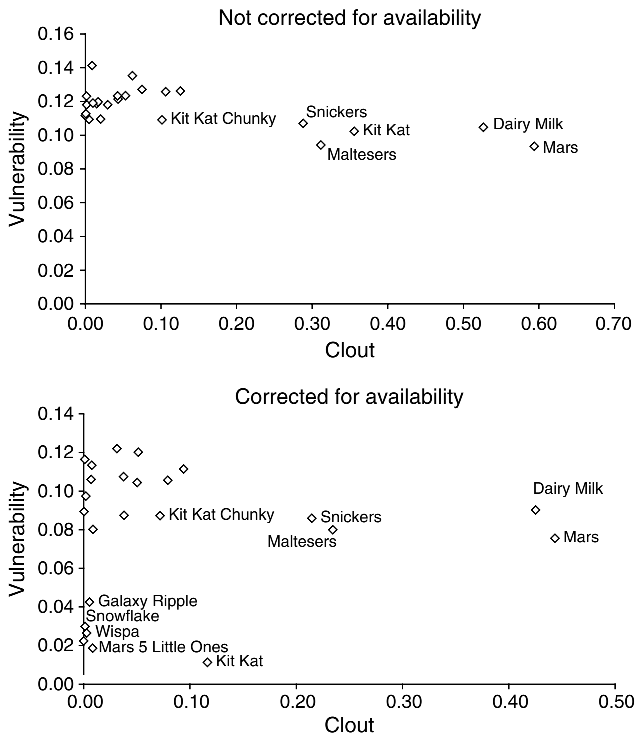
Table 3 Estimates for Product Characteristics and Heterogeneity

Variable format	Estimate	Estimates (standard errors) RC logit	RC logit with availability
Price	Mean	−5.414 (1.121)	−6.061 (0.375)
	Sigma	1.212 (1.185)	2.365 (0.399)
Block	Mean	−2.509 (0.569)	−2.399 (0.447)
	Sigma	0.054 (1.376)	0.279 (1.021)
Indulgence	Mean	−2.828 (0.853)	−3.299 (0.671)
	Sigma	1.063 (0.137)	1.897 (0.177)
Filler	Mean	−3.979 (0.697)	−3.883 (0.548)
	Sigma	0.233 (0.526)	1.122 (0.264)
Bite sized	Mean	−5.042 (0.763)	−5.390 (0.600)
	Sigma	2.459 (0.240)	3.031 (0.189)

Evaluating the competitive landscape is one of the main uses of structural models. To this end, unbiased price coefficients and elasticity measures are very important. If the variation in product availability is ignored, sales variations because of changes in distribution will be captured by the price coefficient (or other environmental factors in more general models). This may result in contamination of the price coefficient by nonprice factors. For instance, if we compare the competitive clout and the vulnerability (Kamakura and Russell 1989)[4] for the products in our data set, we find that the biggest difference occurs for products that have lower levels of availability. In Figure 2, the computed clout and vulnerability are plotted. When estimates are corrected for varying availability, some newly introduced products seem less vulnerable (lower left corner of the lower panel in Figure 2).

[4] Competitive clout is a measure of how other products sales are affected by a change in price of the focal brand, while vulnerability measures the change in sales associated with changes in prices of competing brands.

Figure 2 Competitive Clout and Vulnerability



4. Conclusions

In this paper, we have proposed a straightforward extension to the traditional discrete-choice-based demand estimation method developed in Berry et al. (1995) to incorporate information on product availability. To do that, we assumed that availability is independent across products. We have illustrated our method using data from the UK impulse channel sales of chocolate confectionery products.

Acknowledgments

The authors thank Céline Bonnet and conference participants at the INRA-IDEI *Empirical Industrial Organization of the Food Retailing Industry* for useful discussions. The authors also thank J.-P. Dube, Bart Bronnenberg, and Pradeep Chintagunta for helpful comments, Georg Toufar for making the data available, and the area editor and four anonymous reviewers for comments and suggestions. All errors remain with the authors.

References

Berry, S. 1994. Estimating discrete-choice models of product differentiation. *RAND J. Econom.* **25**(2) 242–262.

Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–889.

Bronnenberg, B. J., C. F. Mela. 2004. Market roll-out and retailer adoption of new brands. *Marketing Sci.* **23**(4) 500–518.

Bronnenberg, B. J., V. Mahajan, W. R. Vanhonacker. 2000. The emergence of market structure in new repeat-purchase categories: The interplay of market share and retailer distribution. *J. Marketing Res.* **37** 16–31.

Campo, K., E. Gijsbrechts, P. Nisol. 2003. The impact of retailer stockouts on whether, how much, and what to buy. *Internat. J. Res. Marketing* **20** 273–286.

Chintagunta, P. K., T. Erdem, P. Rossi, M. Wedel. 2006. Structural modeling in marketing: Review and assessment. *Marketing Sci.* **25**(6) 604–616.

Dolan, R., T. Hayes. 2005. The myths and dangers of ACV. The New England Consulting Group, http://www.brandweek.com/.

Dube, J.-P. 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Sci.* **23**(1) 66–81.

Dube, J.-P., K. Sudhir, A. Ching, G. S. Crawford, M. Draganska, J. T. Fox, W. Hartmann, G. J. Hitsch, V. B. Viard, M. Villas-Boas, N. Vilcassim. 2005. Recent advances in structural econometric modeling: Dynamics, product positioning and entry. *Marketing Lett.* **16**(3–4) 209–224.

Jeuland, A. P. 1979. The interaction effect of preference and availability on brand switching and market share. *Management Sci.* **25**(10) 953–965.

Kadiyali, V., K. Sudhir, V. R. Rao. 2001. Structural analysis of competitive behavior: New empirical industrial organization methods in marketing. *Internat. J. Res. Marketing* **18** 161–186.

Kamakura, W. A., G. J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26** 379–390.

Little, J. D. C. 1998. Integrated measures of sales, merchandising and distribution. *Internat. J. Res. Marketing* **15** 473–485.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.

Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.

Reibstein, D. J., P. W. Farris. 1995. Market share and distribution: A generalization, a speculation, and some implications. *Marketing Sci.* **14**(3, Part 2) G190–G202.

Tenn, S. 2006. Avoiding aggregation bias in demand estimation: A multivariate promotional disaggregation approach. *Quant. Marketing Econom.* **4**(4) 383–405.