Marketing Science

Estimating Demand Heterogeneity Using Aggregated Data: An Application to the Frozen Pizza Category

Paulo Albuquerque, Bart J. Bronnenberg,

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Estimating Demand Heterogeneity Using Aggregated Data: An Application to the Frozen Pizza Category

Paulo Albuquerque
Simon Graduate School of Business, University of Rochester, Rochester, New York 14627,
paulo.albuquerque@simon.rochester.edu

Bart J. Bronnenberg
CentER, Tilburg University, 5000 LE Tilburg, The Netherlands, bart.bronnenberg@uvt.nl

This paper combines different aggregate-level data sets to identify new product demand in consumer packaged goods (CPG) categories. Our approach augments market-level time-series data with widely available summaries of household purchase behavior, i.e., brand penetration and purchase set size data. We show that this augmentation is helpful in the estimation of consumer heterogeneity. For instance, observing a brand with relatively large shares and low penetration typically indicates that preferences are dispersed, with relatively few customers liking the brand a lot. Whereas the combination of share and penetration is informative about heterogeneity with realistic sample sizes, in isolation neither variable may lead to precise estimates of heterogeneity. In addition, other widely available data, e.g., category penetration, is helpful in estimating the size of the total market. Using a large Monte Carlo study, the paper demonstrates the benefits of the proposed approach in estimating model parameters, price elasticities, and brand switching. Empirically, the approach is used to evaluate the launch of a new national brand, DiGiorno, in the frozen pizza category. The new brand is inferred to be very successful at expanding the category, while avoiding cannibalization of existing company share. Using only standard information, i.e., market shares, to estimate the demand model leads, in our data, to poor estimates of the degree of consumer taste variation and of switching to a new brand.

Key words: demand estimation; new products; random coefficients logit model; generalized method of moments
History: Received: November 3, 2006; accepted: March 5, 2008; processed by Guenter Hitsch. Published online in *Articles in Advance* October 24, 2008; revised December 4, 2008.

1. Introduction

Brand switching and new product trial have been studied in the consumer packaged goods (CPG) marketing literature predominantly in the context of the firm's use of price and promotion instruments (e.g., Blattberg and Wisniewski 1989, Carpenter et al. 1988, van Heerde et al. 2003, Van Oest and Franses 2005). Less is known empirically about demand expansion from product innovations, i.e., a new CPG brand,[1] although several papers have called attention to the general topic (e.g., Hauser et al. 2006, Keller and Lehman 2006). Two economic issues have become increasingly important when estimating demand (see, e.g., Chintagunta 2001): endogeneity (usually of prices) and variation in consumer tastes, i.e., consumer heterogeneity. A feasible approach to solving

the endogeneity problem has been suggested by Berry et al. (1995), henceforth BLP, who propose an algorithm to identify the unobserved demand shocks taken into account by manufacturers when setting price. This in turn enables the use of instrumental variables (IV) estimators in determining consumer price effects.

Determining the degree of consumer heterogeneity is important in the context of evaluating new product demand. A common approach to the inference of consumer heterogeneity (e.g., Nevo 2000) is to use aggregate-level time series or cross sections of market share.[2] However, heterogeneity is potentially difficult to estimate using aggregate-level share data. The only information available to identify heterogeneity is the result of discrepancies in observed shares movements and the expected movements predicted by a

---

[1] There is more recent work available on durable goods. Luo et al. (2007) studied the effects of new product introductions on consumer durable goods subject to retailer acceptance criteria. Sriram et al. (2006) study the dynamics in demand for new technology products.

[2] Frequently, the use of individual-level data is obstructed because of lack of availability, high cost, or sparsity and nonrepresentativeness of the data at a local level.

Figure 1 Two Dimensions of Data Used in Estimation

Panel data structure
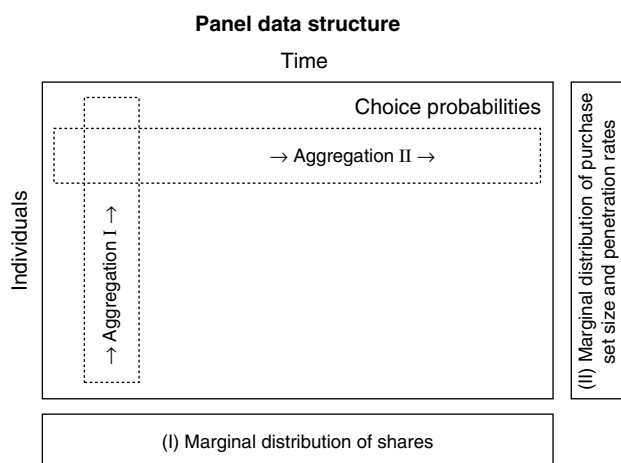
homogeneous model. In cases where these discrepancies are small, heterogeneity parameters will only be weakly identified, if at all (Bodapati and Gupta 2004). Petrin (2002) points out the same predicament, arguing that heterogeneity is identified only if some unusual substitution patterns not captured by the homogeneous model do occur and/or a change in the choice set is present, e.g., the introduction of a new brand.

In this paper, we aim to improve the estimation of demand, and of consumer heterogeneity in particular, using aggregate-level data. Our approach is to combine different sources of aggregated data, or to combine the aggregate-level market share data, with other widely available syndicated CPG data, in this case summary data about consumer purchase behavior. Our proposed additional information is also aggregate level, but it is aggregated in a different direction. Figure 1 illustrates what we mean by this. With individual-level data, the demand data populate an N consumers by T time periods panel, consisting of choices among the incumbent brands and adoption behavior for new brands. As pointed out above, it is common to infer the demand primitives from market shares, which is a summary across consumers.[3] In contrast, we propose to use information from the marginal summaries of both dimensions of the panel by adding data that are aggregations of choice probabilities across time. Using standard sampling theory, these aggregates are still highly accurate even if the individual panel data are too sparse, too costly, or simply not at the disposal of the analyst.

Specifically, we add two summaries of purchase behavior across time. First, we incorporate data on observed purchase set sizes, i.e., the number of unique brands that a consumer switches among in a 12-month period. Second, we use brand and category penetration rates, i.e., the fraction of consumers who buy a particular brand (or any brand from the category) during a 12-month period. We next require that the estimated demand primitives not only match the actual time series of market shares, but also the 12-month summaries of purchase behavior. This improves the identification of preference dispersion (the Results section explains why in more detail) as well as the popularity of the category as a whole (i.e., it helps identify the size of the outside good).

Combining these matching restrictions with the orthogonality restrictions of the standard IV approach (see BLP) can be done using the generalized method of moments (GMM).

In sum, the idea is to identify the demand primitives (including consumer heterogeneity) by placing additional identifying restrictions on them using fundamentally different marginal summaries of the consumer panel data. This process of "triangulation" can be extended in GMM to include other manifestations of the demand primitives and may be particularly applicable to the general CPG industry for which demand data other than market shares are widely available, e.g., from the Information Resources, Inc. (IRI) Factbook.

The combination of multiple sources of information to improve a demand model's accuracy is not new in the context of durable goods. Specifically, the pioneering works by Petrin (2002) and Imbens and Lancaster (1994) show that the inclusion of micro data and consumer surveys helps in the identification of demand primitives such as consumer heterogeneity (see also Berry et al. 2004). Compared to Petrin (2002), we use information on consumer differences in purchase behavior in the estimation instead of demographic consumer characteristics given purchase. This is particularly suitable in a CPG context where such information is reliable and easy to obtain and at the same time informative about demand heterogeneity. Second, we focus on repeat purchase items. This distinction from consumer durable goods deserves some acknowledgment because demand for a new repeat purchase good comes from both trial as well as repeat sales and the inferred degree of preference dispersion strongly affects predictions of both. Third, our approach also differs from past work in that we use the additional data to estimate rather than assume the size of the outside good.[4]

---

[3] We note that the panel may include other dimensions, such as markets (e.g., Nevo 2001) or a larger variety of products (e.g., Berry et al. 1995), in addition to or in lieu of the time dimension. As with time series, demand data at these units of observation still consist of aggregations across consumers.

[4] Recently, Musalem et al. (2008) propose a Bayesian approach to identifying individual-level demand from aggregate data. Their approach requires no additional data, but uses functional form restrictions for identification.

Our intended contributions are as follows. First, using a large scale Monte Carlo study, we show the impact of the augmented data on the quality of demand estimates. We find that adding data about the purchase set size distribution and brand penetration in the market helps us estimate taste variation in the market. Ignoring this information leads to incorrect inferences about brand switching.

Second, aggregate market share data are also uninformative about the size of the outside good, for which separate identification assumptions need to be made. Past studies warn that inferences about substitution and switching behavior are highly dependent on correct estimation of preference heterogeneity (Berry et al. 2004) and the size of the outside good (Nevo 2000). Our simulations show that category penetration helps us in estimating the size of the weekly outside good.

Third, we apply our model to data from the frozen pizza category and focus on evaluating the launch by Kraft of the DiGiorno brand in the Houston, Texas, market. We estimate the relative importance of competitive draw, cannibalization, and category expansion[5] in this category. We find that the new premium-priced DiGiorno brand was very successful at attracting new consumers from outside of the frozen pizza category. In the Houston market, we further find that cannibalization of Kraft's incumbent brands was virtually absent.

The next section presents our demand model. Section 3 describes the estimation algorithm. Section 4 reports on the Monte Carlo study, and §5 discusses the empirical application. We conclude in §6.

2. Model

Our demand model is formulated at the individual level. In each week $t = 1, \ldots, T$, the utility of brand $j = 1, \ldots, J$ for consumer $i = 1, \ldots, N$ is given by the following expression:

$$u_{ijt} = \alpha_{ij} + x_{jt}\beta_i + \xi_{jt} + \epsilon_{ijt}, \qquad (1)$$

where α_{ij} is individual i's preference for brand j, x_{jt} is a K-dimensional row vector of observed marketing mix variables, β_i is a K-dimensional column vector of individual specific marketing mix coefficients, and ξ_{jt} includes demand shocks that are unobserved by the econometrician but considered by consumers

in their purchase decisions and by manufacturers in their pricing decisions.[6]

Consumers are allowed to be heterogenous in their preferences for brands and in their sensitivities to marketing mix variables. For logical consistency, we specify the individual level effects of marketing mix variables to be of their expected sign, using a log-normal random-effects distribution. For example in the case of price effects, β_i is modeled as

$$-\beta_i = \exp(\beta_0 + \sigma_\nu v_i), \quad v_i \sim N(0, 1). \qquad (2)$$

The $[J \times 1]$ vector of brand intercepts $\boldsymbol{\alpha}_i = [\alpha_{i1} \cdots \alpha_{ij} \cdots \alpha_{iJ}]'$ has a multivariate normal distribution with mean $\bar{\boldsymbol{\alpha}} = [\bar{\alpha}_1 \cdots \bar{\alpha}_j \cdots \bar{\alpha}_J]'$ and variance-covariance matrix $\boldsymbol{\Sigma}$ of size $[J \times J]$. In principle, the variance covariance matrix Σ can be fully estimated, but the fact that this would involve the estimation of $J(J+1)/2$ parameters, has lead to the prevalence of more parsimonious specifications. At least two parsimonious specifications for the matrix Σ exist. First, it can be represented as a diagonal matrix with variance terms to be estimated. In this setup, the parameters of the model to be estimated are $\theta = [\bar{\alpha}_j, \beta_0, \sigma]$, where σ is the vector of standard deviations of the brand and marketing mix random effects in the model. Throughout the paper, we will refer to this model as the "diagonal" model.

Second, it can be represented using a factor structure

$$\alpha_{ij} = \bar{\alpha}_j + L_j \omega_i, \qquad (3)$$

with $\omega_i \sim N(0, I)$; see e.g., Chintagunta et al. (2002). In this formulation, L_j is the $[1 \times P]$ vector of coordinates of alternative j in the P-dimensional unobserved attribute space ($P < J$), sometimes interpreted as a perceptual map (Shugan 1987), and ω_i is a $[P \times 1]$ vector of consumer tastes for these attributes. Arraying the J coordinates L_j into a $[J \times P]$ matrix, the distributional assumptions on ω_i imply that $\Sigma = E(L\omega_i\omega_i'L') = LL'$.[7] If P is not too large, a significant reduction in parameters can be obtained relative

---

[5] *Competitive* draw is defined as the fraction of demand for a new product that is caused by consumers switching from competing brands. *Cannibalization* is defined as the fraction of demand that comes from consumers switching from the other brands marketed by the new brand's manufacturer. Finally, *category expansion* is defined as the fraction of demand for a new product that originates from consumers who bought other—indirect—substitutes before.

[6] This model is static in the sense of having constant parameters. We tested a formulation where we allowed the brand positions to change between pre- and post-entry periods (see §5.1 for more details). We found very little variation in the positions of the brands, justifying the more parsimonious model. The stability in brand positions is also found in van Heerde et al. (2004), where only one of the brands' intercepts shows a significant change in this category. Note also that in the latter paper, the variance-covariance of the errors is essentially static, which is equivalent to our fixed-positions approach.

[7] The factor model cannot be estimated without several identification restrictions. Specifically, because of translation invariance, we fix the outside good to be placed in the origin of the attribute space. Because of rotation invariance, we require one alternative to be positioned along the positive horizontal axis. Finally, because of reflection invariance, we restrict the second attribute of the second brand to be positive.

to the free specification of Σ, often without sacrificing too much flexibility of the model. With this setup, the parameters to be estimated are $\theta = [\bar{\alpha}_j, \beta_0, L, \sigma]$, where L are the attribute levels of the products and σ is the vector of standard deviations of the random effects of the marketing mix. Throughout the paper, we will refer to this model as the "factor" model.

The factor model is of interest in our empirical setting for a number of reasons. First, a factor model directly estimates brand similarity in unobserved attributes (e.g., Elrod 1988, Elrod and Keane 1995, Erdem 1996). This property is of interest, especially in the context of DiGiorno's advertising claim that it substitutes with delivery pizza as evidenced by the slogan "It's not delivery, it's DiGiorno!" From this claim, we could expect that DiGiorno substitutes with the outside good, a fact that is directly verifiable from how close DiGiorno is positioned to the outside good in the brand map. The factor model introduces correlation in the unobservable brand characteristics across brands, with relatively few parameters. It thus has the advantage of reducing the number of parameters required to estimate a full (in the sense of nondiagonal) heterogeneity matrix while remaining highly flexible.[8]

There are several reasons to include the mean brand effects, $\bar{\alpha}_j$, into the model. First, it is not certain that observed product characteristics capture all or much of the substitution patterns in the data. In such cases, "fixed effects should be included to improve the fit of the model" (Nevo 2000, p. 536). Second, the random shocks ξ_{jt} may be related to prices. By accounting for brand-specific mean utility components, we also account for possible correlation between prices and the brand-specific mean of unobserved quality. In turn, this has the advantage that we do not need an instrument for this correlation. Last, accounting for the mean alternative-specific utility means that the interpretation of the random shocks ξ_{jt} becomes more precise. Specifically, with the mean utility accounted for, the random shocks ξ_{jt} are zero in expectation at the brand level and represent temporal variability in utility (e.g., because of calendar seasons, or special events such as Superbowl Sunday, etc.). Pricing may depend on such seasonality in a different way than it does on brand differences in unobserved attributes. Thus, by accounting for brand level mean utilities in each market, we can disentangle these two sources of endogeneity, which otherwise would be left confounded.

In practice, consumers can choose among several choice options, the so-called "inside goods," or decide to buy something else (including "nothing") in a given week, the so-called "outside good," which we represent with $j = 0$. Its utility is normalized to ϵ_{i0t} for identification purposes.

Under the assumption that ϵ_{ijt} is drawn from the extreme value distribution, the probability of household i purchasing brand j at time t is given by

$$\mathrm{Pr}_{ijt}(X_t, \xi_t, \theta) = \frac{\exp(\alpha_{ij} + x_{jt}\beta_i + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\alpha_{ik} + x_{kt}\beta_i + \xi_{kt})}. \quad (4)$$

Observed measures of demand, e.g., market share time series, but also purchase set size distributions, and brand penetration data are all different functions or manifestations of these choice probabilities. We use this property in the demand estimation.

3. Demand Estimation

3.1. Overview

We estimate the demand model using the generalized method of moments (Hansen 1982). GMM accommodates combining different sets of information, assigning optimal weights to each piece of data (Imbens and Lancaster 1994), while allowing the use of instrument variables to correct for the correlation that is generally present between price data and unobservable demand shocks (see also Petrin 2002). The possibility to combine multiple data sets and use instrument variables makes the estimation method ideal for our purpose.

We use three different sets of moment conditions. First, we use moments similar to Berry et al. (1995), and Nevo (2001). These moments require that the demand shocks, ξ_{jt}, are orthogonal to a set of instrument variables (to be specified). Second, to estimate the size of the outside good, we define moment restrictions combining the weekly dynamics in category sales and brand penetration rates. Finally, we use a third set of moment restrictions using the brand penetration rates and purchase set size data to aid the identification of taste variation. We now present the details for the implementation of each of these moments.[9]

3.2. The BLP Moments

In empirical studies of demand, the analyst often lacks observation of certain demand primitives that

---

[8] The factor model introduces correlation between alternatives. Recent work by Kayande et al. (2007) has focused instead on correlation between attributes.

[9] We do not model the supply side, i.e., prices, to help estimate the demand parameters. For the observed shelf prices to be informative about the demand parameters in local markets, many assumptions are required about local pricing decisions by national multiproduct firms and about the local category management strategies of retailers. The data offer little or no guidance in making such assumptions and wrong assumptions may deteriorate rather than improve our demand estimates. For instance, in our data, retail prices for the incumbents are similar before and after the launch of DiGiorno. This is at once consistent with the manufacturer charging the same price and the retailer absorbing the shock in wholesale prices (e.g., Nelson et al. 1992).

are observed and used by the manufacturer as inputs to the determination of price. This causes correlation between prices and unobserved attributes ξ_{jt}, and generally leads to biases in the estimates of the demand parameters. Past literature has provided evidence of this so-called endogeneity bias when using store-level data (Chintagunta 2001, Villas-Boas and Winer 1999).

To account for the endogeneity of price, the usual approach is to rely on IV and impose an orthogonality condition with the unobserved demand shocks ξ_{jt}. Berry et al. (1995) have proposed an algorithm to estimate the ξ_{jt}. In this algorithm, the indirect utility function $\alpha_{ij} + x_{jt}\beta_i + \xi_{jt}$ is divided in an individual part, in our case $\mu_{ijt} = L_j\omega_i + x_{jt}\beta_i$, and a mean utility for brand j at time t, $\delta_{jt} = \bar{\alpha}_j + \xi_{jt}$.[10] Next, given an initial value δ_{jt}^0 and a set of parameter values, the following expression is iterated until it converges[11]

$$\delta_{jt}^{n+1} = \delta_{jt}^n + \ln(s_{jt}) - \ln(\hat{s}_{jt}(\delta_{jt}^n, \theta)), \qquad (5)$$

where $\hat{s}_{jt}$ is the expectation of the choice probabilities in Equation (4) taken over the distribution of individuals i. Using $\phi(.)$ to denote the PDF of the normal distribution,

$$\hat{s}_{jt} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pr_{ijt}(X_t, \xi_t, \theta)\phi(v)\phi(\omega)\,\partial v\,\partial\omega. \quad (6)$$

Furthermore, s_{jt} is the actual share, and n counts the iterations in the BLP contraction mapping of Equation (5).

The shares s_{jt} are not actually observed. Instead, what is observed is the share among the inside goods; i.e., the conditional shares $\tilde{s}_{jt} = s_{jt}/(1 - s_{0t})$. In practice, the translation from the observations $\tilde{s}_{jt}$ to the shares s_{jt} is made by an assumption about the total size of the market, and thereby an assumption about s_{0t}. If no satisfactory assumption about s_{0t} is readily available, our approach allows for an estimation of it. It does so by replacing s_{jt} in the estimation with the share among the inside goods, $\tilde{s}_{jt}$, which is data, multiplied by 1 minus the share for the outside good (which we estimate; see the next subsection).

$$s_{jt} = \tilde{s}_{jt} \times (1 - s_{0t}). \qquad (7)$$

Given our additional moment restrictions below, this suffices for identification.[12]

In the empirical section, we define a set of instruments Z_{jt} that correlate with the potentially endogenous variables X_{jt} but not with the unobserved demand shocks ξ_{jt}. This orthogonality can be exploited to construct an IV estimator, as proposed by BLP. Specifically, using the vector of instruments Z_{jt} and the scalar shock ξ_{jt}, we write the "BLP moments" as

$$G_1(\theta): \mathrm{E}[\xi_{jt}(\theta) \otimes Z_{jt}] = 0, \qquad (8)$$

where the expectation is taken over products and time.

3.3. The Outside Good Moments
The size of the outside good is usually not observed, especially not in a CPG context, where purchase incidence can fluctuate seasonally or through the use of promotion instruments. Nevo (2000) notes that there are generally two assumptions in determining the size of the outside good. First, one should choose a variable to which the total size of the market is assumed proportional, and second, one should choose the value of the proportionality factor. Nevo (2000) also observes that these choices influence conclusions about demand systems and substitution effects. In this paper, we propose to set the weekly share of the inside goods proportional to the total weekly expenditure across all categories (CE_t) in a market[13] and we then estimate the—nonstructural—proportionality or scaling factor using data.

The proportionality factor is determined as follows. In Equation (7), we replace s_{0t} with

$$\hat{s}_{0t} = 1 - \lambda \times CE_t, \qquad (9)$$

and we define a moment that chooses the scaling factor λ such that the model is consistent with observed brand and category penetration.[14] Brand and category penetration identify λ, because—as in Nevo's observation—different estimates $\hat{s}_{0t}$ for s_{0t} will generate different parameters θ that in turn imply different penetration rates. Thus in estimation, the structural parameters θ are a function λ, i.e., write $\theta(\lambda)$. Extant papers have the same conditionality, but our

---

[10] Because the β_i have a log normal distribution, we do not factor out the population mean as a linear parameter, to be included in the δ_{jt}, as Nevo (2000) does.

[11] Convergence is obtained $|\delta_{jt}^{n+1} - \delta_{jt}^n| < \varepsilon$, for $\forall \delta_{jt}$, with ε very small, i.e., $10E - 10$, in this study.

[12] This modification is of course in and by itself close to the current practice in estimating demand models. Indeed, it may be realized that current practice also uses Equation (7); however, that it makes a priori choice about the quantity s_{0t} that is contained in it. In our case, we allow for an estimate of this quantity.

[13] Thus, category volume is measured as the share of CE_t among all categories scanned in a given market. Alternatively, one can make the joint share of the inside goods proportional to observed category sales. However, total recorded category sales in our data are subject to dynamics in the IRI store sample. This makes that the dynamics of the frozen pizza share of total store sales is better at capturing the dynamics of the joint sales of the inside goods.

[14] A notational distinction between the parameters θ and λ is made on the grounds that the former are the structural parameters of the demand system, while the latter is a nonstructural scaling constant that translates data about category size into "data" about the outside good. To have a meaningful interpretation of λ as the combined size of the inside goods, we normalize, without any loss in generality, CE_t to have a mean of 1.

paper differs from those in that we estimate λ rather than assume it.

To evaluate the moment restriction, we need to compute the annual category penetration rate implied by the model. This can be done using the choice probabilities in Equation (4). Namely, for each simulated household i (resulting from a draw of v_i and ω_i), brand j, and week t, the model predicts a choice probability $\Pr_{ijt}(X_t, \xi_t, \theta)$. Further, define consumer i's purchase set $\{C_i\}$ as containing all brands bought by i at least once during a year. The probability that i chooses only the outside good over a single year is[15]

$$\Pr_i(\{\varnothing\}) = \prod_{t=T-51}^{T} \Pr_{i0t}(X_t, \xi_t, \theta). \qquad (10)$$

This compound probability is smooth in θ and λ. We determine λ by requiring that the population mean of this probability, $E[\Pr_i(\{\varnothing\})]$, is equal to one minus the observed category penetration rate, π_c.

Equations similar to (10) can be formulated for the model's predictions about brand penetration. The individual-level probability that brand j is chosen at least once within a 52-week period is equal to one minus the joint probability that the brand was never chosen in the 52-week period.

$$\Pr_i(j \in \{C_i\}) = 1 - \prod_{t=T-51}^{T} (1 - \Pr_{ijt}(X_t, \xi_t, \theta)). \qquad (11)$$

Observed brand penetration π_j is the expectation of this quantity across individuals, i.e., the population mean, $E[\Pr(j \in \{C_i\})]$, of the probability that the purchase set contains j (which can be computed in estimation through simulation).

Arraying these $J+1$ conditions, we write the "outside good" moments as

$$G_2(\theta, \lambda): \ E \begin{bmatrix} \Pr_i(\{\varnothing\}) \\ \Pr_i(1 \in \{C_i\}) \\ \vdots \\ \Pr_i(J \in \{C_i\}) \end{bmatrix} = \begin{bmatrix} 1 - \pi_c \\ \pi_1 \\ \vdots \\ \pi_J \end{bmatrix}, \qquad (12)$$

where the expectation is taken over individuals i.

3.4. The Heterogeneity Moments
In addition to the brand penetration data, we use the distribution of purchase set sizes, S_i, to help further

identify the dispersion of preferences and, importantly in the evaluation of a new product introduction, the degree of switching in the frozen pizza category. Our data cover the empirical distribution of the purchase set size for frozen pizza across households, $\Pr(S_i = 0)$, $\Pr(S_i = 1)$, $\Pr(S_i = 2)$, etc., in different regions in the United States. For example, in the West South Central Census division, 29% of households buy zero frozen pizza brands in a year (therefore category penetration is 71%), 25% of households buy only one unique brand, 19% switch between two brands, 14% switch among three brands, and 7% switch among four brands (the remaining 7% of households switch among more than four brands).

We recursively compute the predicted purchase set size distribution of the model from the implied choice probabilities, $\Pr_{ijt}(X_t, \xi_t, \theta)$ in Equation (4). As an example, we provide details on the model's predictions for $\Pr(S_i = 1)$ and $\Pr(S_i = 2)$.

Start with the joint probability that a weekly observed consumer buys brand j, nothing, or combinations thereof over the course of a year,

$$H_{ij} = \prod_{t=T-51}^{T} [\Pr_{ijt}(X_t, \xi_t, \theta) + \Pr_{i0t}(X_t, \xi_t, \theta)]. \qquad (13)$$

This probability covers all purchase histories that combine any number of purchases of j with any number of purchases of the outside good. Therefore (using the notation in Equation (10)),

$$\Pr_i(\{j\}) = H_{ij} - \Pr_i(\{\varnothing\}) \qquad (14)$$

is the probability that the purchase set is $\{j\}$ in a given year. Finally, the probability that the consumer has a purchase set size of exactly one, is equal to the summation of $\Pr_i(\{j\})$ across choice options $j \neq 0$.

$$\Pr(S_i = 1) = \sum_{j=1}^{J} \Pr_i(\{j\}). \qquad (15)$$

Next, $\Pr(S_i = 2)$ can be computed starting with the probability that the consumer purchases j, k, nothing, or combinations thereof for a year,

$$H_{ijk} = \prod_{t=T-51}^{T} \big[\Pr_{ijt}(X_t, \xi_t, \theta) + \Pr_{ikt}(X_t, \xi_t, \theta)$$
$$ + \Pr_{i0t}(X_t, \xi_t, \theta) \big]. \qquad (16)$$

This probability covers all purchase histories involving j, k, and the outside good. The probability $\Pr_i(\{j, k\})$ that the consumer's purchase set is $\{j, k\}$, i.e., that the purchase set contains at least one j and one k but no other brands besides the outside good is then (using Equation (14))

$$\Pr_i(\{j, k\}) = H_{ijk} - \Pr_i(\{j\}) - \Pr_i(\{k\}) - \Pr_i(\{\varnothing\}). \qquad (17)$$

---

[15] Note that we use the last 52 weeks (from $T-51$ to T) of observations for the computation of the penetration rate. During these weeks there are no new brand introductions and the choice set is stable. The moments in this subsection and the next are computed using postlaunch data; i.e., the predictions and data are matched only over the final 52 weeks of postlaunch data. Thus, we do not apply the same category penetration rate to the period before and after DiGiorno.

As a final step, the probability that for a randomly selected individual a purchase set of exactly size two is observed equals the sum of $\Pr_i(\{j, k\})$ across all unique combinations of j and $k \neq 0$.

$$\Pr(S_i = 2) = \sum_{j=1}^{J} \sum_{k=j+1}^{J} \Pr_i(\{j, k\}). \qquad (18)$$

The probabilities $\Pr(S_i = 3)$ and $\Pr(S_i = 4)$ are recursively computed in a similar fashion. We match the predictions of $\Pr(S_i = s)$ for $s = 1, \ldots, 4$ to the actual data. Write the population values for the fractions $\Pr(S_i = s)$ as $\mathcal{F}_s$. Then, the final set of moments can be written as

$$G_3(\theta): \mathrm{E}[\Pr(S_i = s)] = \mathcal{F}_s, \quad s = \{1, \ldots, 4\}, \qquad (19)$$

where the expectation is again taken over households.

With data on $\mathcal{F}_s$, this set of moments ensures that the model parameters are chosen such that the implied amount of switching given prices, promotion, etc., matches the switching in the frozen pizza category observed during the introduction of DiGiorno.

3.5. Objective Function and Simulation
The objective function combines the three sets of moments previously described:

$$G(\theta) = \begin{bmatrix} G_1(\theta) \\ G_2(\theta) \\ G_3(\theta) \end{bmatrix}. \qquad (20)$$

To compute the expectations in $G_1(\theta)$, $G_2(\theta)$, and $G_3(\theta)$ we need to use simulation. For instance, the expectation in Equation (19),

$$\mathrm{E}[\Pr(S_i = s)] = \iint \Pr(S_i = s)\phi(v)\phi(\omega)\,\partial v\,\partial \omega, \qquad (21)$$

cannot be computed analytically but must be approximated. To this end, we can use the pseudo-panel of (v_i, ω_i) draws that is also used for the approximation of the market share integrals in $G_1(\theta)$.

$$\mathrm{E}[\Pr(S_i = s)] \approx \frac{1}{N} \sum_{i=1}^{N} \Pr(S_i = s \mid \theta, X_t, v_i, \omega_i). \qquad (22)$$

This approximation is again smooth in the parameters θ. The same can be done for the expectation in Equation (12).

Next, we use these approximations in a two-step GMM estimator (Hansen 1982, Petrin 2002):

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \big(\widehat{G}(\theta)' W(\tilde{\theta})' W(\tilde{\theta}) \widehat{G}(\theta)\big), \qquad (23)$$

where $\widehat{G}(\theta)$ is the sample analogue of $G(\theta)$ and $W(\tilde{\theta})$ is a weight matrix consisting of an estimate of the "square root" of the inverse of the variance-

covariance matrix of the moments, obtained using $\tilde{\theta}$, a preliminary consistent estimate of θ.

For the first set of moments, $G_1(\theta)$, the weight matrix is given by

$$W_1(\tilde{\theta})' W_1(\tilde{\theta}) = \left[\frac{1}{T^2} \sum_{t=1}^{T} g_{1t}(\tilde{\theta}) \cdot g_{1t}(\tilde{\theta})' \right]^{-1}, \qquad (24)$$

where $g_{1t}(\tilde{\theta})$ are the moment values for each time period.

Under the assumptions of the model, the variability of the second and third set of moments originates from alternative realizations of the random demand shocks ξ_{jt}. Consequently, we can compute the variance of the moments by evaluating how different draws from ξ_{jt} affect $G_2(\theta)$, and $G_3(\theta)$. We do so by sampling with replacement from the empirical distribution of $\xi_{jt}(\tilde{\theta})$ and computing the empirical value of $G_2(\tilde{\theta})$, and $G_3(\tilde{\theta})$ using this sample of "ξ-draws." By replicating this process a number of times, we obtain a sample of moment values from which the variance in the moments can be computed directly.[16] The inverse of this matrix is the desired weight matrix $W_2(\tilde{\theta})' W_2(\tilde{\theta})$.

Finally, we used for the complete weight matrix $W(\tilde{\theta})' W(\tilde{\theta})$ the block-diagonal combination of the two parts defined above (see Petrin 2002 for a similar approach).

3.6. Computing Local Switching
In our empirical example, we evaluate the introduction of DiGiorno. To obtain the switching from incumbent brands to DiGiorno, we compare two scenarios: the actual scenario, where DiGiorno was introduced in the market, and an alternative counterfactual case, where we remove DiGiorno from the market by setting its utility to $-\infty$. We then compute the difference of shares of incumbent brands in the two scenarios. The idea behind this method is to identify which brand would have kept the share that was transferred to the introduced brand.

Formally, brand switching is computed using the following expression:

$$\Delta s_{jt} = \frac{1}{N} \sum_{i=1}^{N} [\Pr_{ijt}(\theta, X_t, \text{DiGiorno in})$$
$$- \Pr_{ijt}(\theta, X_t, \text{DiGiorno out})]$$
$$j = 1, \ldots, J, \quad j \neq \text{DiGiorno}$$
$$\forall t \text{ after DiGiorno's entry.} \qquad (25)$$

---

[16] This variance measure translates the variance in demand shocks to variance of the moments. Note that it is easy to account for additional measurement error. For instance, if it is known that the penetration measures are only accurate up to plus or minus 1%, one can add this noise as a diagonal variance matrix. Finally, simulation error is neglible and can be made arbitrarily small by increasing the number of simulation draws. We tested alternative measures of variance, with similar results.

Under the assumptions of the model, this measure is less than or equal to 0 (incumbent brands will not gain share from the introduction of DiGiorno) and larger than minus the share of the incumbent brands prior to the launch of DiGiorno.[17]

4. Monte Carlo Simulation

4.1. Data Generation and Experiment Design

To assess the impact of the additional moments on the estimates of the demand system we conduct a numerical experiment. Because the diagonal model is the most widely used in empirical work, we focus on this model in the experiment. We generate data according to the utility model (1) and probability model (4) in the paper. This creates an $N \times J \times T$ table of choice probabilities for N simulated households, J brands, and T time periods. For the generation of the data, we choose $N = 5{,}000$ households, $J = 6$ brands, $T = 104$ weeks. During the first 52 weeks, five brands are present. A single new brand is launched in week 53. The actual prices and promotion data from a U.S. market (Chicago) are used in the generation of the choice data. Consumers are generated with different tastes for each of the brands and with different price sensitivities. The variances of the random effects are brand specific. The values for the data generating parameters of the demand model are set at realistic values, similar to those obtained empirically using data for the market of Chicago.

The instruments are defined as follows. For each combination of year and brand we include an intercept. That is to say, we require ξ_{jt} to have a mean of zero prelaunch and postlaunch of the new brand. We use prices in three "far away" markets (see the empirical section) and promotion variables as further instruments, as well as the square of these marketing mix instruments.

To approximate the demand integrals, we use 500 pseudo-households.[18] Starting values for the parameters in the GMM estimation are computed using a nonlinear least squares (NLS) estimator. This estimator minimizes the squared deviations between the model predictions and the data jointly, across both marginals of Figure 1. To combine the fit in time series

with the fit of the purchase set size/penetration data, a weighted sum is used that makes both components equally important. This NLS estimator converges fast but does not account for price endogeneity. It is therefore only used to obtain preliminary values for the parameter estimates prior to using GMM.

The Monte Carlo study contains three "conditions," each representing an estimation regime. In the first, we use all available information but keep the outside good fixed (at the correct value). In the second condition, we again use all available information but now the size of the outside good is estimated along with other demand parameters. Finally, in the third condition, we assume the correct size of the outside good as in condition 1, but we ignore the additional information about purchase set sizes and penetration rates, and instead estimate the model using the market share time series only.

For each replication of the experiment, we kept the generated data, the household draws, and the starting values of the demand parameters constant across the three conditions. This facilitates comparison of the results across conditions.

4.2. Results

We ran the experiment 100 times and saved several measures of model fit and the point estimates of the demand system for analysis. We first comment on model fit. Table 1 shows that all experimental conditions have essentially the same mean absolute deviation (MAD) in market share fit (i.e., the difference between actual share and predicted share computed at the expectation of the demand shocks, i.e., with $\xi_{jt} = 0$). We conclude that all estimation regimes lead to similar fit of the time series of market shares. The exact value of the temporal MAD reflects the variance in the unobserved demand shocks.

Whereas in conditions 1 and 2 we observe a good fit between the model and the data on purchase set sizes and brand penetration, the fit is poor in condition 3, where we only use the time-series market share data to estimate the model parameters. It is not surprising that the augmented information leads to better fit of the purchase set size and brand penetration data. What is surprising is that poor fit in

---

[17] Another option is to compare the market shares of the incumbent brands pre- and postlaunch by DiGiorno. However, this contrast is not purely attributable to the launch of DiGiorno, as many exogenous things may have changed (random demand shocks, promotion variables, etc). In addition, this contrast tells us little about the change in size of the outside good, which needs to be inferred through the use of a model.

[18] In addition to using 500 draws in the simulation, we tested models with 250, 500, and even 1,000 draws, with no apparent difference in estimation results. To be conservative, in the empirical example, we use 1,000 draws.

Table 1   Fit Measures in the Data Experiment

Experimental condition	1	2	3
Information	Augmented		Standard
Share of inside good	Known	Estimated	Known
MAD temporal	0.0244	0.0244	0.0258
MAD purchase set	0.0030	0.0028	0.0975
MAD brand penetration	0.0053	0.0053	0.1150

Figure 2   Three Key Parameters Across the Conditions of the Numerical Experiment (*Nobs* = 100)
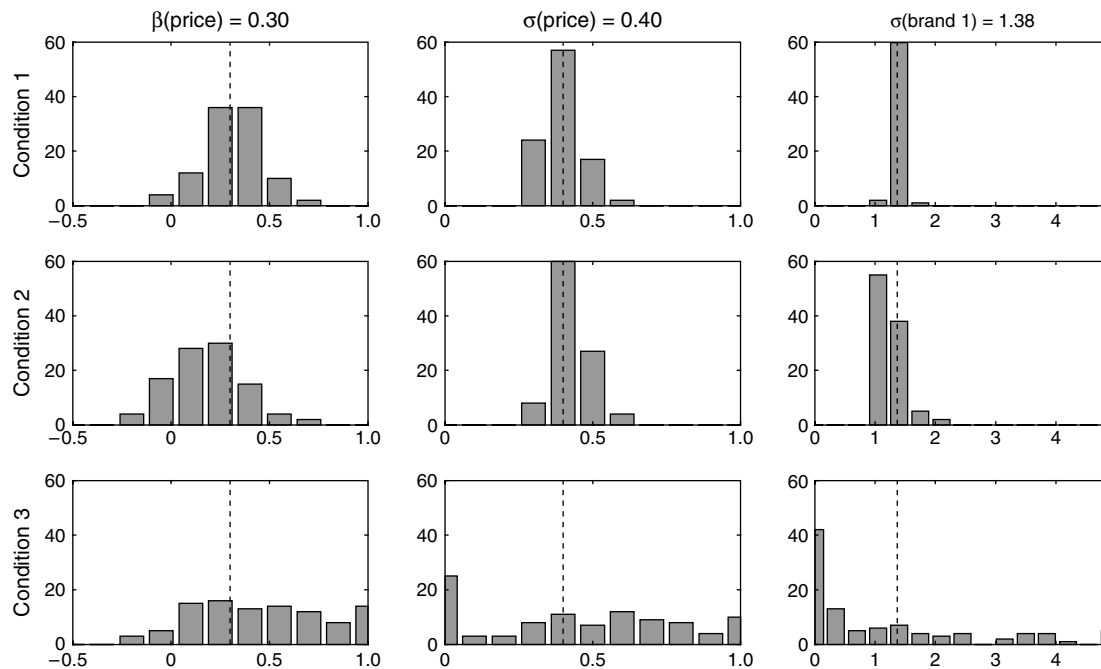
terms of purchase set sizes, and brand penetration—as evidenced in condition 3—does not affect how well the model fits the market shares. We therefore conclude that, under the random effects logit model, market share data alone are not very informative about important demand characteristics such as intensity of brand switching and brand penetration.

We next discuss the distribution of several key demand parameters in the experiment. Figure 2 shows the histograms of the point estimates for the location parameter (β_0) in the lognormal distribution of price coefficients (the data generating value is 0.30),[19] standard deviation in household price responses (0.40), and standard deviation of the random effects for brand 1 (1.38). First, we can conclude that the heterogeneity parameters are well recovered using the augmented information (conditions 1 and 2). Focusing on condition 2, even when we concurrently estimate the size of the outside good, we can still recover heterogeneity in price responses and brand preferences nearly as well as having exact knowledge of the size of the outside good, albeit that the heterogeneity in brand preferences is inferred with somewhat more variance.

Contrasting these findings with condition 3, we observe that in absence of the extra information, the heterogeneity parameters are poorly recovered. In many instances, the variance parameters either

tend to 0 or take on large values (even in some cases to an estimation upper bound that for practical reasons was set to 10 in the experiment).

This pattern generalizes to the other parameters of the model. Table 2 summarizes the other results of the numerical experiment and reports the mean and standard deviation (across replications) of the point

---

[19] Recall that the random price effects in our model are negative with a lognormal distribution with mean β_0 and variance σ_ν^2 (see Equation (2)).

Table 2   Estimation Results in Simulation Study

		Mean estimated value (*Nobs* = 100) (standard deviation across replications)		
	True value	Condition 1	Condition 2	Condition 3
Marketing mix				
Price	0.300	0.315 (0.157)	0.185 (0.183)	0.504 (0.415)
Display	1.300	1.267 (0.171)	1.324 (0.148)	1.454 (0.310)
Feature	0.100	0.091 (0.093)	0.117 (0.089)	0.131 (0.110)
Brand intercepts				
Brand 1	−1.700	−1.596 (0.586)	−2.014 (0.591)	−1.874 (2.422)
Brand 2	−2.500	−2.393 (0.545)	−2.727 (0.533)	−2.414 (1.475)
Brand 3	−1.600	−1.493 (0.578)	−1.916 (0.594)	−1.294 (1.652)
Brand 4	−4.100	−4.028 (0.561)	−4.211 (0.483)	−4.276 (3.262)
Brand 5	−3.800	−3.714 (0.446)	−3.794 (0.365)	−4.266 (3.310)
Brand 6	−1.000	−0.889 (0.631)	−1.354 (0.654)	−3.736 (5.505)
Standard deviation (heterogeneity)				
Brand 1	1.378	1.381 (0.076)	1.251 (0.191)	1.271 (1.821)
Brand 2	1.414	1.401 (0.133)	1.239 (0.205)	1.190 (1.315)
Brand 3	1.549	1.558 (0.070)	1.427 (0.177)	1.040 (1.452)
Brand 4	1.483	1.518 (0.160)	1.217 (0.362)	1.228 (1.828)
Brand 5	1.183	1.190 (0.126)	0.582 (0.553)	1.186 (1.840)
Brand 6	1.049	1.050 (0.083)	0.907 (0.195)	2.659 (3.577)
Price	0.400	0.400 (0.064)	0.424 (0.063)	0.475 (0.389)
Outside good				
Scale	0.099	0.099[a]	0.088 (0.016)	0.099[a]

[a]Not estimated but fixed at actual value.

Figure 3   **Three Demand Characteristics Across the Conditions of the Numerical Experiment (*Nobs* = 100)**
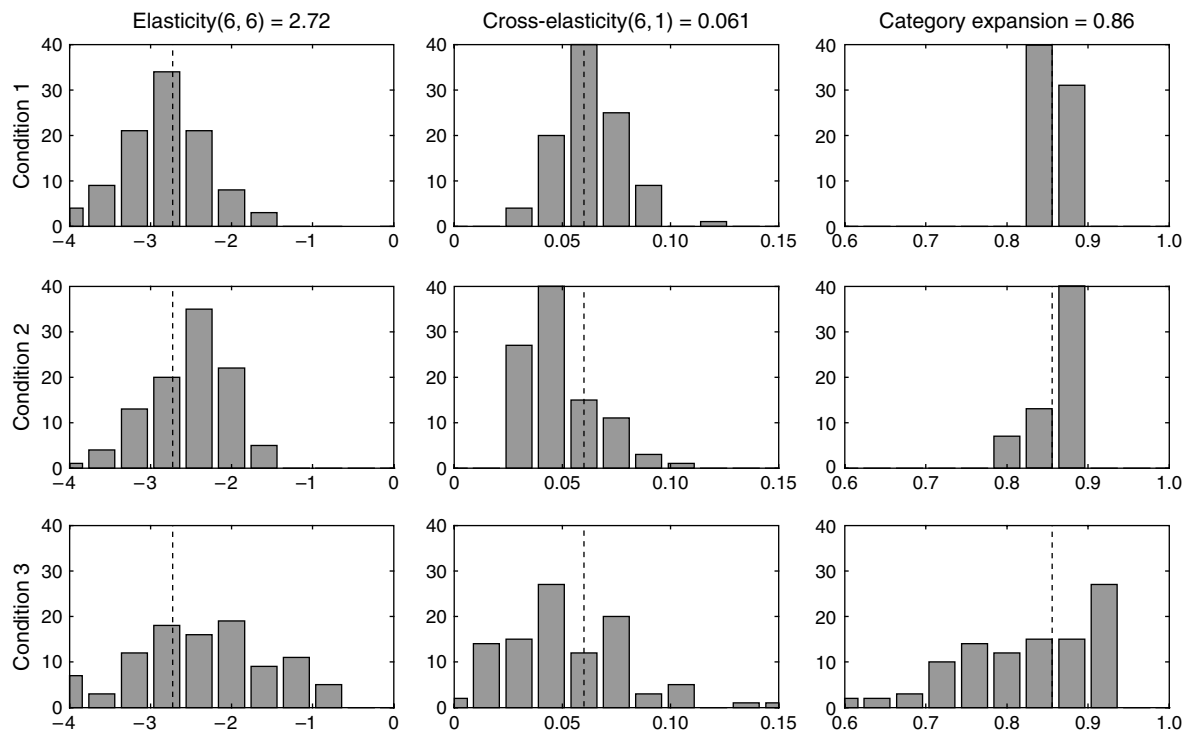
estimates of the model parameters. The main result of the analysis is that the heterogeneity parameters are subject to large inference errors when we use the standard information to estimate the model. This can be observed from the column in the table that is labeled "condition 3." For instance, the taste variation in brand 2 is estimated to be 1.190 on average (true value is 1.414), but the standard error of that estimate is 1.315. Note that these results are likely conservative because we assumed the correct size of the outside good, and because we used the same preliminary values for the parameters across all conditions, and these preliminary values were computed using the augmented information.

We further note from the column that is labeled "condition 2" in the table, that the outside good is estimable. That is, the true value of λ in Equation (9) is 0.099 (i.e., the combined size of the inside goods is 9.9% in the data experiment), and the estimate for this quantity is 0.088 with a standard error of 0.016. Contrasting this with the column that is labeled "condition 1" reveals that the estimation of the outside good comes at the expense of some efficiency in the estimates of taste variation. There is also some underestimation of the degree of consumer taste variation, especially in the case of brand 5. The differences between the results in conditions 1 and 2 appear to be caused by the limited sample size and suggest that when a reasonable choice for the size of the outside good is available, such information is still valuable in the context of our augmented information.

We conclude that the additional moments involving purchase set size and brand penetration used in condition 1 (and in condition 2) substantially improve the efficiency of estimates of taste variation relative to condition 3.

To see how differences in the demand parameters translate into differences in demand characteristics, we post-processed the 100 replications and computed three different types of demand characteristics. Specifically, we report on (1) the own price elasticity of the new brand, (2) the cross-price elasticities between this brand and brand 1, and (3) the fraction of demand for the new product that comes from the outside good. Figure 3 shows the histograms of the point estimates of these quantities. It also shows the values of these quantities at the data generating parameters (again with a hatched line). The results in condition 1 show that (cross) elasticities and category expansion are all centered around their actual values. When we estimate the outside good (condition 2), rather than assuming it, the variance of the cross-elasticity estimate becomes higher and the mode shifts slightly towards zero, but the estimates for own elasticity are virtually identical. Also, the implied fraction of demand that is drawn from the outside good in conditions 1 and 2 are very similar to its actual value. Thus, we conclude that the share and the augmented information are collectively useful in identifying (cross) elasticity and category expansion.

However, in condition 3, where we use only market share data, estimates of both the cross- as well as own elasticities have much more variance across replications. In addition, the estimates of category expansion display large variation and are too small in many cases. This is due in large part to the frequent over-estimation of price heterogeneity (see Figure 2) which creates a large tail of price-sensitive consumers who choose the outside good and do not want to try the new premium-priced brand.

To conclude, our simulation results support that readily available data on purchase set size and brand penetration (1) improves the fit of demand models in other dimensions than the time series, (2) improves the estimates of the demand parameters, and (3) helps estimate demand characteristics such as elasticities and the origins of new brand demand. The simulation also shows that taste variation is poorly identified from the market share data by the orthogonality conditions in GMM.[20]

5. Empirical Analysis

5.1. Data

Our empirical analysis covers the frozen pizza category and within that category we focus on evaluating the launch of DiGiorno. Frozen pizza has become one of the most important categories among frozen food, accounting for about 19% of its sales (Bronnenberg and Mela 2004, van Heerde et al. 2004). According to industry experts and manufacturers, it represents almost 20% of the total pizza business, with delivery pizza being its main competitor outside of the category (*Pizza Marketing Quarterly* 2000). Between 1993 and 1995, the years preceding our analysis, the category was characterized by slow growth, with dollar sales marginally increasing from $1.6 to $1.7 billion. In 1995, Kraft launched a new brand into the market, DiGiorno. In late 1996, Schwan's followed by launching Freschetta. Both brands included a new feature, self-rising crust, which was considered a major development in the category. Combined with strong advertising, DiGiorno's introduction led to a fast increase in sales of frozen pizza with a sustained annual growth rate of approximately 12% through 1999 (Holcomb 2000).

Kraft and Schwan's Food Company are the dominant players in the frozen pizza category and each compete with multiple brands. Kraft's brands include DiGiorno, Tombstone, and Jack's, while Schwan's

Table 3 Evolution of Average Shares for the Main Brands in the Frozen Pizza Category in Each of the Years in the Data Set

	1995	1996	1997	1998	1999
DiGiorno	0.01	0.04	0.10	0.13	0.13
Jack's	0.09	0.09	0.07	0.07	0.08
Red Baron	0.12	0.12	0.11	0.12	0.13
Tombstone	0.22	0.21	0.19	0.18	0.17
Tony's	0.12	0.12	0.10	0.09	0.08
Totino's	0.11	0.11	0.11	0.11	0.10
Other brands	0.33	0.31	0.31	0.30	0.31

owns Tony's and Red Baron. Another national brand in this category is Totino's, which is owned by Pillsbury. Our analysis of the introduction of DiGiorno will focus mainly on these six brands, which capture about 70% of the national volume of the category.[21] All of them, except Jack's, are available nationally. Jack's distribution is limited to markets in the Northwest and Midwest region of the country, but the brand has a large share in those markets.

In hopes of avoiding cannibalization of its existing brands, Tombstone and Jack's, Kraft exploited DiGiorno's rising crust attribute in its marketing. Specifically, because rising crust was associated with fresh baked or hand-tossed pizza, Kraft positioned DiGiorno as substitute for delivery pizza instead of traditional frozen pizza.

Average annual shares from 1995 to 1999 for the main brands are presented in Table 3. Nationally, the dynamics in DiGiorno's share reflect a roll-out that took three years. In our data, DiGiorno captured about 13% of the U.S. frozen pizza market by 1999.[22]

Although we have data and estimation results for several U.S. cities, we report on the launch of DiGiorno in Houston, Texas. Figure 4 shows the evolution in market shares of three of the brands on that market. To better illustrate the overall trends in the data, the figure shows 13-week moving windows. We can see that the market share for DiGiorno reaches approximately 12% and that the share builds quickly after local launch. The market share for Red Baron drops modestly from 26% to 23%; the market share for Tony's also drops modestly from 14% to 12%. The remainder of DiGiorno's share comes from other brands.

Our empirical analysis integrates three different data sets. The first data set covers market-level sales volume, price, local feature advertising and promotional display use in Houston. The data, covering 260 weeks from January 1995 to December 1999, are constructed by aggregating over a sample of stores

---

[20] Our model is intended for capturing CPG data, e.g., time series of relatively few brands. It is not the same model as in BLP, who had (1) a much larger cross section of products, (2) no product-level fixed effects, (3) included the supply-side equations to help estimate the demand parameters.

[21] Freschetta is not included in our analysis of the introduction of DiGiorno because it was introduced later.

[22] The same figure is independently reported in Holcomb (2000).

Figure 4   Evolution of Market Shares in Houston (Smoothed with a 13-Week Moving Window)
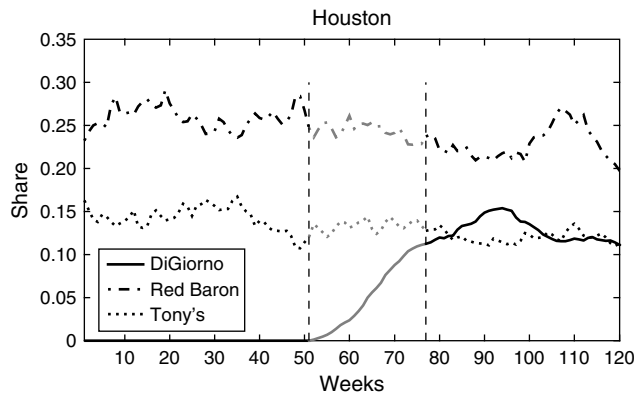
in the Houston market. We use volume sales to compute the market shares of the inside goods. For our empirical analysis, we do not use a 26-week window immediately following the launch of DiGiorno. The data in this window display dynamics of postlaunch sales that often reflect local depth of distribution more than demand (see, e.g., Bronnenberg and Mela 2004). We are primarily interested in consumer substitution patterns that explain the differences between pre- and postlaunch market shares given distribution. For the empirical analysis, we therefore censor the 26-week period after the launch of DiGiorno (see Figure 4). Thus, the market share data are represented by two time series, one representing the situation before and the second the situation after the launch of DiGiorno, both for a period of 52 weeks. Note that two years of market share data is typical of sample sizes for store level CPG data.

The second data set consists of weekly data on the local size of the frozen pizza category as a fraction of total store volume. These data are informative about the dynamics in category volume (the total size of the "inside" goods) in the Houston market.

The third set of data (ACNielsen 2001) consists of summary statistics of purchase behavior using Nielsen's HomeScan panel. From these data, we have access to the local distribution of purchase set sizes, i.e., the percentage of consumers that buy 0, 1, 2, 3, 4, or more unique brands during one year. We have these data for the Census division to which Houston belongs[23] and for the year 2004. We also have access to annual brand penetration levels for each Census division, for the years 2000 to 2003, measuring the

percentage of people that have purchased a given brand of pizza at least once during a year.[24]

5.2.   Estimation Details
A number of estimation details warrant discussion. To approximate the demand integrals, we use 1,000 pseudo-households. Using a set of preliminary estimates,[25] we next estimate the parameters in two stages. In the first stage, consistent initial estimates of all the parameters are obtained and the optimal weight matrix in GMM is computed. In the second stage, this optimal weight matrix is used to compute the final parameter estimates and their standard errors.

We use the following instruments. First, as in the experiment, we include a dummy variable for each brand and "period," where a period is defined as a full year prior to the launch of DiGiorno or a full year after the launch of DiGiorno (see Figure 4). Using these as instruments implies that the model fits the mean share for each j, before and after the launch of DiGiorno. In turn, this makes the model correctly fit the share adjustments among the incumbents to the new brand.[26] For price instruments, we use the prices of three far-away markets for the brands (see, e.g., Nevo 2001). Far-away markets are identified by sorting markets according to increasing distance to the

---

[23] There are nine U.S. Census divisions: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. For further definitions, see http://www.census.gov/geo/www/us_regdiv.pdf.

[24] A point of concern is that these data were collected a few years later than the sales data. This is a limitation of the data and not of the approach or of its practical scope. We tested the robustness of our findings by assuming that the penetration and purchase set size data are observed with error and adjusting the weight matrix in Equation (23), accordingly. Our results do not change substantively.

[25] We determined preliminary parameter values for the factor model by jointly minimizing the sum of squared errors in the time series and in the consumer purchase data (penetration and purchase set size). Because the factor model can be subject to local minima, we did this 25 times from randomly drawn initial values and retained the estimates of the best fitting model. In the presence of endogeneity, this "NLS" estimate for price effects is not consistent, but this procedure still provides a useful set of starting values for the heterogeneity parameters of the model in the context of GMM estimation. We did this separately for the "augmented information" case, and for the "standard information" case. This procedure precedes and does not replace the first stage in the two stage GMM estimation.

[26] Note that we estimate as many brand intercepts $\bar{\alpha}_j$ in the model as we have brands. Our choice for instruments perhaps evokes the question why we did not choose to have "double" intercepts in the model for each incumbent brand, one pre- and one post launch of DiGiorno. This is technically feasible. However, it is theoretically not appealing to specify a shift in utility common to all consumers because of the introduction of a new brand, especially if this is done to compensate for the fact that the implied substitution of the specified demand model does not predict the correct market adjustments. Rather, we model substitution through the random effects structure and leverage the idea that the pre- to postlaunch adjustments in share are informative about how brand utilities correlate among consumers. Thus, rather than allowing for jumps in brand constants, we let the introduction of the new brand help us identify the correlation structure of utilities.

market being analyzed, and as instruments, we choose prices from markets ranked 11th to 13th, subject to the new brand being present in those markets. Given the assumed absence of endogeneity in display and feature activity, we also used display and feature as instruments. Finally, we also used the squares of the price and promotion instruments.[27]

Without any loss in generality, we estimate the factor model with polar coordinates rather than Cartesian coordinates in the attribute space. This facilitates the interpretation of the factor model in terms of variances and correlations in brand preferences. For instance, with a two-dimensional representation of the attribute space, instead of estimating positions $[L_{1j}, L_{2j}]$ in attribute space, we estimate the length of the attribute vector σ_j and angle with the origin ϕ_j. Note that this implies a one-to-one transformation $[L_{1j}, L_{2j}] = [\sigma_j \cos(\phi_j), \sigma_j \sin(\phi_j)]$. With this formulation, element j, k of the variance covariance matrix LL' is equal to $L_j L_k' = \sigma_j \sigma_k (\cos(\phi_j) \cos(\phi_k) + \sin(\phi_j) \sin(\phi_k))$. From the so-called composite argument property of trigonometric functions, this expression equals the following form:

$$L_j L_k' = \sigma_j \sigma_k \cos(\phi_j - \phi_k). \quad (26)$$

The convenient aspect of this formulation is that it factors out variance and correlation terms. For instance, the jth diagonal element of the covariance matrix LL' can be obtained from this formula by setting $j = k$, and is equal to $\sigma_j \sigma_j \cos(\phi_j - \phi_j) = \sigma_j^2$. Furthermore, the term $\cos(\phi_j - \phi_k)$ in Equation (26) is easily recognized as the correlation between preferences for products j and k among the consumer population. Thus, in this formulation, if two brands are on perpendicular rays onto the origin, their preferences are unrelated in the consumer population. On the other hand, if they are on the same ray onto the origin, their correlation is $+1$.

A final detail is that to keep the discussion compact, we discuss the factor model only. The results with the diagonal model are substantively similar.

We present our estimation results in the following order. First, we report on the fit of the model. Second, we discuss several structural parameter estimates. Finally, we discuss the implied origins of demand for the newly launched brand, DiGiorno.

5.3. Model Fit

As in the Monte Carlo study, we first briefly evaluate how well the model explains market shares, the distribution of the purchase set size, and brand

Figure 5 In-Sample Fit of the Model in the Market of Houston, Texas
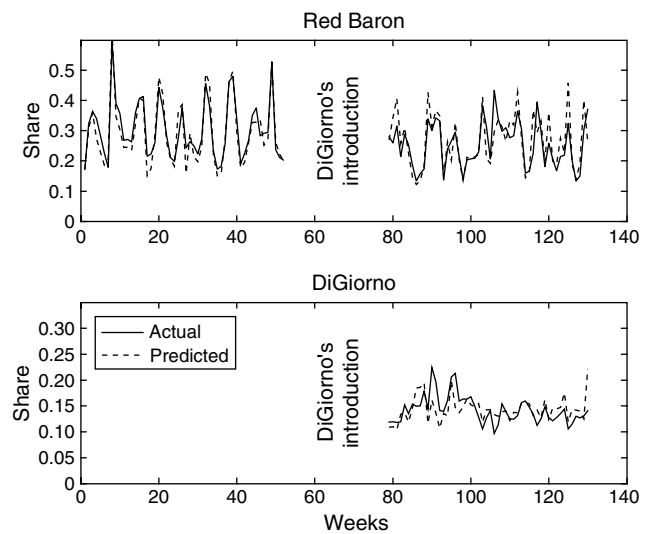
penetration. To evaluate the improvement stemming from the additional moments, we compare our proposed model (containing the augmented information) with the model in which the outside good and heterogeneity moments are not included in the estimation (the "standard" model). In the standard model, we use the estimated size of the outside good from the augmented model as data.[28]

To illustrate the findings, we discuss the case of Houston, Texas. Figure 5 displays the time series of actual and estimated shares for two brands in this market.[29] It is clear that the proposed model fits the temporal variation in market shares as well as the average market shares very well. As in the data experiment, the standard model also does well in recovering the market share time series.

Next, Table 4 shows that the demand parameters obtained from the augmented model correctly predict the actual purchase set sizes observed in the market. In contrast, the standard model does not, even when we use the information about the outside good borrowed from the augmented model. Indeed, the last column of the table shows that the standard model strongly overestimates single- and dual-brand loyalty (purchase set size of 1 and 2). At the same time it underestimates the fraction of households that switch among many products.

The augmented model also fits the brand penetration data very well. However, the estimates from the standard model are far from the actual observa-

---

[27] In the augmented information case there are more moment conditions than there are parameters. Therefore, we took the average of the price instruments and the average of the square of the price instruments. In the standard information case, the separate price instruments are necessary.

[28] Because the standard model uses the outside good estimates from the augmented model, the standard model is likely to perform better than it would with an ad hoc assumption about the outside good. The contrast in relative improvement of model fit from the extra moments is therefore once more likely to be conservative.

[29] The shares are estimated excluding the demand unobservables ξ_{jt} and the error term ϵ_{ijt}, as these are not observed by the analyst.

Table 4   In-Sample Fit of Purchase Set Size and Penetration Rates

		Factor model	
	Actual	Augmented model	Standard model
Purchase set size			
0 brands	0.289	0.322	0.246
1 brand	0.249	0.275	0.341
2 brands	0.187	0.199	0.309
3 brands	0.135	0.138	0.086
4 brands	0.074	0.063	0.019
Brand penetration rates			
Tombstone	0.181	0.175	0.456
Red Baron	0.408	0.433	0.241
Tony's	0.186	0.205	0.202
Totino's	0.261	0.266	0.254
DiGiorno	0.264	0.273	0.139

tions. For instance, the standard model implies that Tombstone is bought at least once a year by 45.6% of the households in the market, whereas its actual penetration is less than half of that estimate. The standard model implies that Red Baron has a penetration of 24.1%, whereas its actual penetration is 40.8%. The standard model implies that the demand for DiGiorno comes from too few customers, i.e., 13.9% instead of 26.4% of the population (because the model fits the shares very well, these customers also buy the brand too frequently). These implications of the standard model disagree with the additional data. Thus, as in the Monte Carlo study, the underlying problem with the standard model is that the heterogeneity parameters are not well identified (see also our discussion of the model parameters momentarily). We thus conclude that, empirically, the market share data by itself appear to be relatively uninformative under the orthogonality conditions $E(\xi_{jt} \otimes Z_{jt} = 0)$ about brand penetration and trial (in the case of DiGiorno).

5.4. Structural Parameters and Brand Perceptions

We now report on the estimates of the demand parameters of the factor model. Table 5 presents the estimates and standard errors for the factor model, using the augmented information and the standard information. Note that the price parameter is positive but the price effect is negative (see Equation (2)).

First, a big difference between the two sets of estimates is that the standard errors of the parameters become very large in the case of model estimates with standard information. Using the standard information, the intercepts and the brand heterogeneity parameters are inferred to be strongly negatively correlated.[30] In contrast, adding the purchase set size and

[30] We have used additional instruments and experimented with adding different transformations of existing instruments, e.g., log or 3rd power, but this does not reduce the standard errors in a meaningful way.

Table 5   Estimates of the Demand Parameters in Houston

	Factor model	
	Augmented	Standard
Marketing mix		
Price	0.369 (0.357)	0.156 (1.157)
Display	0.874 (0.264)	1.069 (2.554)
Feature	0.221 (0.231)	0.400 (0.935)
Intercepts		
Tombstone	−5.159 (0.992)	−1.469 (5.360)
Red Baron	−0.928 (1.592)	−4.095 (14.672)
Tony's	−3.369 (1.601)	−4.481 (3.832)
Totino's	−4.430 (0.763)	−4.904 (25.538)
DiGiorno	−1.928 (1.707)	−6.164 (34.372)
Standard deviation (= length of attribute vector)		
Tombstone	4.572 (1.366)	0.697 (12.867)
Red Baron	2.423 (0.394)	3.642 (7.650)
Tony's	3.501 (0.598)	3.510 (1.967)
Totino's	4.156 (0.896)	3.607 (14.782)
DiGiorno	2.138 (0.442)	4.334 (18.431)
Angle attribute vector		
Tombstone	0.000 (0.000)	0.000 (0.000)
Red Baron	0.292 (0.120)	0.721 (1.450)
Tony's	0.285 (0.024)	0.707 (1.556)
Totino's	1.990 (0.145)	0.072 (5.832)
DiGiorno	0.711 (0.153)	1.143 (0.720)
Sigma		
Price	0.118 (1.003)	0.377 (1.291)
Scale	0.127 (0.040)	0.127 (0.000)

Note. Standard errors in parentheses.

brand penetration data greatly reduces the standard errors of the estimates.

Second, the standard deviations of the random brand effects are all estimated to be different from zero in the augmented information case. In contrast, none of the standard deviations are significant using the standard information.

Third, there is an interesting observation about the estimated location of brands in the unobserved attribute space. Recall that the outside good occupies the origin in this attribute space. The standard deviations reported in the table are the radius of the location of each of the brands (see the section on estimation details). Thus, brands with small taste variances are "located" closer to the outside good, than brands with larger taste variances. Such brands are closer substitutes to the outside good. We find that in the Houston market, using the augmented information, the brands Tombstone and Totino's are subject to large taste variations, whereas the DiGiorno brand has the smallest degree of taste variation (relative to the outside good). We therefore infer that of all available brands, it is perceived as closest to the outside good. This resonates well with the advertising slogan that the brand has used for years, in which it positions itself as a brand close to the outside good (in this case delivery).

Table 6 Estimates of the Switching to DiGiorno, for the Market of Houston Using the Augmented Model and the Standard Model, as a Percentage of DiGiorno's Share

	Factor model	
	Augmented	Standard
Cannibalization	0.000	0.009
Draw	0.072	0.081
Expansion	0.928	0.910
Elasticity	−3.069	−1.684
Trial	0.273	0.139

Table 7 Purchase Behavior and Preference Heterogeneity

	Heterogeneity	Penetration	Share	Share penetration
Tombstone	4.572	0.181	0.147	−0.033
Red Baron	2.423	0.408	0.254	−0.154
Tony's	3.501	0.186	0.135	−0.051
Totino's	4.156	0.261	0.322	0.061
DiGiorno	2.138	0.264	0.142	−0.121

In contrast, using the standard information, the brand is positioned far away from the outside good.

Multiplied by π, the "angles" reported in Table 5 constitute the angle between the brand's radius and the horizontal axis. For instance, in the perceptual map, the DiGiorno brand is estimated to be located at an angle of 0.711π relative to the horizontal axis. With a radius of 2.138, it is the only brand in this location of the perceptual map, suggesting that DiGiorno is perceived as a unique brand. Indeed, DiGiorno has unique characteristics, e.g., rising crust, not found in other brands, which in turn lends credibility to the inferred perceptual map (and thus to the associated variance covariance structure).

The estimation of the outside good share is fundamental to the process of accurately identifying switching between alternatives because in many cases category expansion plays an important role in generating demand for a new brand, especially one that offers unique product features. Using our approach, we find that our estimate of the scale parameter λ is equal to 0.127. This implies an estimated weekly share of the outside good of $1 - 0.127 = 0.873$.

5.5. The Impact of the Extra Information
Table 6 lists the results of a counter factual experiment, where we removed DiGiorno from the market, and evaluated how shares readjust to this policy. The model with the augmented information indicates (and the model with the standard information confirms) that, in the Houston market, the demand for DiGiorno originates almost entirely from the outside good and that sales for DiGiorno are almost fully incremental. The cannibalization from Tombstone is estimated to be almost zero. Also the draw from competing products is estimated to be modest. Indeed, from a new product launch perspective, this constitutes a desirable scenario.

The market for frozen pizza is fairly price sensitive. Own elasticity for the new brand is estimated at −3.069. We note that the elasticities are markedly lower (−1.684) when estimated using the standard

information, which is consistent with the higher inferred degree of price heterogeneity.[31]

Finally, Table 7 contrasts share and penetration data to provide an intuition for why the additional information yields better estimates of heterogeneity. Specifically, it lists the estimated brand-level variances (heterogeneity) for the main brands. It also displays the observed penetration rates and market share data, along with the difference of the share and penetration data. Note that Totino's has a high share relative to its penetration, whereas say, Red Baron has the exact opposite. Thus, relative to Red Baron, Totino's appeals to fewer customers. But, given its high share, Totino's customers have a higher utility for Totino's than Red Baron's customers have for Red Baron. This means that the preferences for Totino's must be more dispersed in the population than Red Baron's. Indeed, our heterogeneity estimates are consistent with this. The correlation between the heterogeneity estimates and the difference between share and penetration (second and fifth columns of Table 7) is 0.83. Thus, the combination of brand share and brand penetration data leads to a strong correspondence with the estimated degree of heterogeneity. In contrast, brand share itself does not correlate with the brand-level standard deviations (0.12). In other words, average shares are not informative about heterogeneity but, theoretically and practically, the difference (or ratio) between shares and penetration is.

In conclusion, when estimated using the augmented information, the factor model offers intuitive estimates of the brand positioning of DiGiorno. It also offers estimates of taste variations for the brands in the market. These empirical estimates strongly correlate with the difference of share and penetration data. The estimates have substantially smaller standard errors using the augmented information than using the standard information. In sum, we believe that the additional information helps create better and more intuitive estimates of brand heterogeneity.

6. Conclusion
In this study, we analyzed demand for a new CPG product. Researchers often have a recurring dilemma

---

[31] Generally, only the right tail of the price coefficient distribution is in the market. Those that are too price sensitive will be loyal to the outside good.

in the estimation of demand. On the one hand, large samples of individual purchases for all markets/stores under analysis contain very rich data but are still difficult to obtain. On the other hand, market-level data are easier and cheaper to obtain, but are considerably less informative about individual-level behavior, as details are lost in aggregation. Our analysis offers a feasible solution to this conundrum by combining multiple sources of information, all readily available to the marketing manager or the interested analyst. Specifically, we propose to augment the time series of sales or market share, which are summaries of the individual-level data across households, with statistics of consumer purchase behavior that summarize aspects of the individual-level data across time periods. Technically, the paper aims to offer an improvement to the estimation of heterogeneity using aggregated data within an IV/GMM framework.

We propose more specifically that measures of brand penetration and the household distribution of purchase set size contain useful information for the inference of taste variation in combination with market share data. The purchase set size is a measure of the intensity of brand switching. The degree to which consumers switch across many or few brands depends on how similar or dispersed their preferences for the brands are. For instance, consumers who have smaller purchase set sizes, tend to have preferences that are more dispersed. In addition to purchase set size measures, taste variation is also revealed through the combination of brand penetration and market share data. For instance, we expect that brands with low penetration and high share generally have higher degrees of taste variation among consumers than brands with high penetration but low share. Using a Monte Carlo experiment, we have shown that these two additional sources of aggregate data lead to large improvements in the inference of individual-level preference dispersion in a GMM estimation approach. In contrast, using time-series data of market shares alone gives poor estimates of preference dispersion.

Another methodological contribution of our approach is that we estimate the overall size of the outside good by relating popularity of the outside good to observed local consumer tendencies to stay out of the market.

Taken together, our approach is helpful in identifying the three sources of market share of a new brand—cannibalization, competitive draw, and category expansion—through the use of easy-to-obtain market-level data.

Substantively, this paper analyzed the launch of a new CPG brand in the frozen pizza category. From company interviews and from its well-known media campaign, we know Kraft was focusing on the pizza delivery market as one of their main targets. Our estimates of the introduction of DiGiorno confirm that the outside good was the main source of DiGiorno's demand.

Finally, because we can estimate our model using data from a single market, we believe that in practical terms our model is helpful to managers in evaluating the impact of new product introductions in local markets. To the extent that this impact is spatially dependent across markets, our model can be used in phased national roll-outs, such as the one used by Kraft to launch DiGiorno, to forecast new product switching at a preentry stage, based on postentry data from nearby markets.

Acknowledgments

References

ACNielsen. 2001. Consumer insight: New product introductions. Consumer and Market Trends Report, ACNielsen, Chicago.

Berry, S. T., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.

Berry, S. T., J. Levinsohn, A. Pakes. 2004. Differentiated products demand systems from a combination of micro and macro data: The new car market. *J. Political Econom.* **112**(1) 68–105.

Blattberg, R., K. Wisniewski. 1989. Priced-induced patterns of competition. *Marketing Sci.* **8**(4) 291–309.

Bodapati, A., S. Gupta. 2004. The recoverability of segmentation structure from store-level aggregate data. *J. Marketing Res.* **41**(3) 351–364.

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

Carpenter, G. S., L. Cooper, D. M. Hanssens, D. Midgley. 1988. Modeling asymmetric competition. *Marketing Sci.* **7**(4) 393–412.

Chintagunta, P. K. 2001. Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Sci.* **20**(4) 442–56.

Chintagunta, P. K., J.-P. Dubé, V. Singh. 2002. Market structure across stores: An application of a random coefficients logit model with store level data. *Econom. Models Marketing* **16** 191–221.

Elrod, T. 1988. Choice map: Inferring a product market map from panel data. *Marketing Sci.* **7**(1) 21–40.

Elrod, T., M. P. Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *J. Marketing Res.* **32**(1) 1–16.

Erdem, T. 1996. A dynamic analysis of market structure based on panel data. *Marketing Sci.* **15**(4) 359–378.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4) 1029–1054.

Hauser, J. R., G. J. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for *Marketing Science*. *Marketing Sci.* **25**(6) 740–759.

Holcomb, R. 2000. Food industry overview: Frozen Pizza. Report FAPC-103, Food and Agricultural Products Center, Oklahoma State University, Stillwater.

Imbens, G. W., T. Lancaster. 1994. Combining micro and macro data in microeconometric models. *Rev. Econom. Studies* **61**(4) 655–680.

Kayande, U., J. H. Roberts, G. L. Lilien, D. K. H. Fong. 2007. Mapping the bounds of incoherence: How far can you go and how does it affect your brand? *Marketing Sci.* **26**(4) 504–513.

Keller, K. L., D. Lehman. 2006. Brands and branding: Research findings and future priorities. *Marketing Sci.* **25**(6) 740–759.

Luo, L., P. K. Kannan, B. T. Ratchford. 2007. New product development under channel acceptance. *Marketing Sci.* **26**(2) 149–163.

Musalem, A., E. T. Bradlow, J. S. Raju. 2008. Who's got the coupon? Estimating consumer preferences and coupon usage from aggregate information. *J. Marketing Res.* **45**(6) 715–730.

Nelson, P., J. J. Siegfried, J. Howell. 1992. A simultaneous equations model of coffee brand pricing and advertising. *Rev. Econom. Statist.* **74**(1) 54–63.

Nevo, A. 2000. A practitioner's guide to estimation of random-coefficients logit models of demand. *J. Econom. Management Strategy* **9**(4) 513–548.

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

Petrin, A. 2002. Quantifying the benefits of new products: The case of the Minivan. *J. Political Econom.* **110**(4) 705–729.

Pizza Marketing Quarterly. 2000. Rising crust pizza: Should you worry? (Winter) 16–18, http://www.pmq.com/digiorno.shtml.

Shugan, S. M. 1987. Estimating brand positioning maps using supermarket scanning data. *J. Marketing Res.* **24**(1) 1–18.

Sriram, S., P. K. Chintagunta, R. Neelamegham. 2006. Effects of brand preference, product attributes, and marketing mix variables in technology product markets. *Marketing Sci.* **25**(5) 440–456.

van Heerde, H., S. Gupta, D. R. Wittink. 2003. Is 75% of the sales promotion bump due to brand switching? No, only 33% is. *J. Marketing Res.* **40**(4) 481–491.

van Heerde, H., C. F. Mela, P. Manchanda. 2004. The dynamic effect of innovation on market structure. *J. Marketing Res.* **41**(2) 166–183.

Van Oest, R., P. H. Franses. 2005. Which brands gain share from which brands? Inference from store-level scanner data. *Quant. Marketing Econom.* **3**(3) 281–304.

Villas-Boas, J. M., R. Winer. 1999. Endogeneity in brand choice models. *Management Sci.* **45**(10) 1324–1338.