Measuring Consumer Switching to a New Brand of Consumer Packaged Goods*

Paulo Albuquerque† Bart J. Bronnenberg‡

14 August 2007

*We appreciate comments from Dan Ackerberg, Andrew Ainslie, Anand Bodapati, Charles Corbett, Jean-Pierre Dubé, Mike Duffy, Carl Mela, Sanjog Mishra, and Garrett Sonnier. We also acknowledge comments made by seminar participants at Dartmouth College, Duke University, The Invitational Choice Symposium at the Wharton School, Singapore Management University, University of Florida, University of Rochester, University of Science and Technology Hong Kong, University of Texas at Austin, University of Tilburg, and Washington University St. Louis. We thank IRI and AC Nielsen for their generous contributions to the data sets used in this paper.

†Assistant Professor at the Simon Graduate School of Business, University of Rochester, paulo.albuquerque@simon.rochester.edu. Financial support from the Portuguese Foundation for Science and Technology is gratefully acknowledged.

‡Professor at Tilburg University, Tilburg, the Netherlands, b.j.bronnenberg@uvt.nl. Financial support from NSF grant SES-0644767 is gratefully acknowledged.

Measuring Consumer Switching to a New Brand of Consumer Packaged Goods

Abstract

This paper studies local demand for new brands of consumer packaged goods (CPG) and the identification of its underlying sources. We show that the standard choice-based aggregate-level demand model provides poor estimates of the degree of consumer taste variation and of switching to a new brand. We introduce an estimation procedure to avoid these problems and then study new product demand across a national sample of local markets. Our approach augments the market-level time series with widely available summaries of household switching behavior and brand penetration data. This syndicated data has the added benefit that it allows us to estimate the size of the total market rather than assume it. Using a large Monte Carlo study, the paper demonstrates the benefits of our 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 brand share. The study further suggests that substitution patterns to the new national brand are highly regional and not national.

Keywords: new products, random coefficients logit model, national and local branding

1 Introduction

New products are very pervasive in consumer packaged goods (CPG) industries (e.g., Kahn and McAlister 1997; MacLeod 1995). AC Nielsen (2001) reports that there were more than 30,000 new products introduced in CPG industries in 2001 alone. Whereas the majority of these introductions concerns me-too products and line extensions, many mark the launch of more innovative products. New or recently launched products are reported to account for a disproportionate share of profits (Buzzell and Gale 1987), and the prevalence of new product launches is increasing over time (AC Nielsen 2001; Kahn and McAlister 1997). Therefore, there is little disagreement about the long-term importance of successful new product introductions, especially in mature CPG industries that are often characterized by eroded profit-margins and heavy use of price-promotions. To estimate demand for a new brand, especially in an existing category, a manager needs to know with which existing brand(s) it substitutes, ideally in each of several separate local markets.[1] Accordingly, the aim of this paper is to empirically estimate local demand for a new CPG brand and to explore the geographic differences in new product substitution patterns.

Brand switching and new product trial has been studied in the CPG marketing literature mainly 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, Gupta and Wittink 2003; Van Oest and Franses 2005). Yet surprisingly little is known empirically about consumer switching due to the firm's use of product innovations, i.e., a new CPG brand.[2] A common approach to demand estimation (e.g., Nevo 2000) is to use time series (or cross-sections) of market share data to infer the mean and the dispersion (heterogeneity) of consumer tastes for a product and its characteristics. In practice, aggregate level data is used for this task. In our case, we are also confined to aggregate data because to evaluate the local demand for a new product, individual level data is simply not available or too sparse.[3] Aggregate data may work well to identify mean preferences, but it is not clear that it is informative about the dispersion of preferences across the consumer population (see also Bodapati and Gupta, 2004; Petrin, 2002). Further, these 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

[1] With this terminology we mean markets that are demand-separated, i.e., where markets are far enough apart that consumers do not travel between them to benefit from lower prices.

[2] There is more recent work available in durable goods. Luo, Kannan, and Ratchford (2007) studied the effects of new product introductions in consumer durable goods subject to retailer acceptance criteria. Sriram, Chintagunta, and Neelamegham (2006) study the dynamics in demand for new technology products.

[3] Individual consumer panels exist in the U.S., but these panels are too small to reliably estimate local switching among many brands. Therefore, the sellers of the data do not generally provide these data at a disaggregate/local level to firms.

about substitution and switching behavior are highly dependent on correct estimation of preference heterogeneity (Berry, Levinsohn, and Pakes 2004) and the size of the outside good (Nevo 2000). Thus the aggregate nature of the data presents a fundamental problem in identifying the taste heterogeneity in the market and in identifying the sources of new product demand.

We aim to redress this problem by collecting and adding data that are informative about variances in consumer tastes and the market size to the estimation problem. We combine these data in estimation with market share time series, which in turn are informative about the local population mean of preferences, using the Generalized Method of Moments (GMM).

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 (the number of unique brands a consumer switches among) distribution and brand penetration in the market helps us estimate taste variation in the market, and that adding data about the local category penetration to the local time series of market shares allows us to empirically identify the size of the outside good. Ignoring this information leads to incorrect inferences about brand switching.

Second, we estimate our model using data from the Frozen Pizza category and focus on the national launch by Kraft of the DiGiorno brand, and we estimate the relative importance of competitive draw, cannibalization, and category expansion[4] 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. Generally, cannibalization of Kraft's incumbent brands was minor, even in some markets where Kraft already had strong market share.

Third, we show that the demand decomposition for DiGiorno is highly local. Therefore brand substitution is not common across local markets, not even across those markets with the same competitive set.

The next section discusses our general approach in more detail. Section 3 presents our demand model. We describe both the Frozen Pizza industry and the data used to estimate the model in section 4. Section 5 describes the estimation algorithm. Section 6 reports on the Monte Carlo study, and section 7 discusses the results. We conclude in section 8.

[4]Competitive draw is defined as the fraction of demand for a new product that is caused by consumers switching away 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.

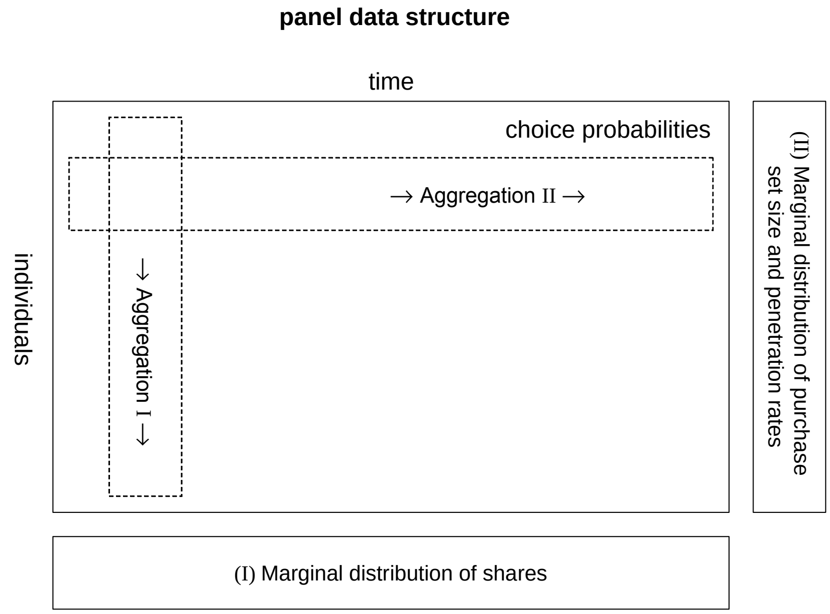

Figure 1: Two dimensions of data used in estimation

2 Our approach

Two economic issues have become increasingly important when estimating demand models (see, e.g., Chintagunta 2001): endogeneity (usually of prices) and consumer heterogeneity. A feasible approach to solving the endogeneity problem has been suggested by Berry, Levinsohn and Pakes (1995), henceforth BLP, who propose an algorithm to identify the unobserved demand shocks that are taken into account by manufacturers when setting price. This in turn enables the use of instrumental variables (IV) estimators in determining consumer price effects.

Heterogeneity is difficult to estimate using aggregate-level share data. Indeed, the only information available to identify heterogeneity is the result of discrepancies in observed shares movements and the expected movements predicted by a 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, defending 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. We show that even when such conditions are met, heterogeneity in preferences, the rate of brand switching, and trial of the new brand are all difficult to estimate empirically from the time-series of market shares alone.

As stated above, our approach maintains the use of market-level data about shares and the marketing mix, but offsets some of the shortcomings of these data with syndicated summary data about consumer purchase behavior. This additional information is also aggregate-level information 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 the new brand. But because this individual-level information is too sparse and too expensive, we propose to use readily available information from the marginal summaries of both dimensions of the panel data (e.g., aggregates across time and consumers) rather than just one.[5] Using standard sampling theory, these summaries or aggregates are still highly accurate even if the panel data are sparse.

Specifically, we add two sources of consumer data on purchase patterns. First, to distinguish between varying degrees of consumer switching, we incorporate data on observed purchase set sizes and brand penetration. We ask that the brand switching predictions obtained from the estimated demand primitives are consistent with observed purchase set sizes and brand level penetration. This matching turns out to identify preference dispersion (the Results section explains why in more detail). Second, to distinguish between varying degrees of category popularity, we require that the fraction of consumers that buy from the category as predicted by the demand primitives is consistent with category penetration and with the dynamics in category sales. This matching helps identify the number of consumers outside the category. The combination of these restrictions with the orthogonality restrictions of the standard IV approach (see BLP) allows us to (better) identify demand parameters. Thus our approach amounts to placing more identifying restrictions on the same set of demand primitives by using data from fundamentally different marginal distributions of the consumer panel data that are nonetheless all widely available. This process of "triangulation" can be extended in a GMM algorithm to include other manifestations of the demand primitives and is particularly applicable to the general CPG industry of which demand data other than market shares are widely available, e.g., from the 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 (e.g., Berry, Levinsohn, and Pakes 2004; Petrin 2002). We add to this literature in several ways. First, we use information on consumer differences in purchase behavior instead of demographic consumer characteristics given purchase (Petrin 2002) to

[5] 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, Levinsohn, and Pakes 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.

the estimation. 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, and correctly identifying the degree of preference dispersion is fundamental to determine the amount of switching that occurs to a new brand. 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.

3 Model

Our demand model is formulated at the individual level, for each market separately. To avoid cluttered notation, we however suppress the market subscript. The model is the same in each market, but the parameter values and the data are market specific.

In each week $t = 1, ..., T$, the utility of brand $j = 1, ..., J$ for consumer $i = 1, ..., N$ is given by the following expression:

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

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 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. Finally, ϵ_{ijt} is an i.i.d. stochastic term with a Type-I Extreme Value distribution.[6]

Consumers are allowed to be heterogenous in their preferences for brands and in their sensitivities to marketing mix variables. 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 $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_j \cdots \alpha_J]'$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The matrix Σ of size $[J \times J]$ can be represented using a factor structure $L_j\omega_i$, with $\omega_i \sim N(0, I)$, see e.g., Chintagunta, Dubé, and Singh (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 $[L \times P]$ matrix, the distributional assumptions on ω_i imply that $\Sigma = E\left(L\omega_i\omega_i' L'\right) = LL'$.[7]

[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 the Data section 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, Mela, and Manchanda (JMR, 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 withouth several identification restrictions. Specifically, because of transla-

We model the variance-covariance matrix Σ using a factor model for several 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. Second, the factor model introduces correlation in the unobservable brand characteristics across brands, with relatively few parameters. It thus presents the advantage of reducing the number of parameters required to estimate a full (in the sense of non-diagonal) heterogeneity matrix while remaining highly flexible.[8]

The introduction of the mean brand effects, α_j, introduces brand dummies into the model. There are several reasons to include such brand level fixed effects. 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). Second, the random shocks ξ_{jt} may be related to prices. By accounting for the alternative 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 seasonal variability in demand (e.g. calendar seasons, or special events such as Superbowl Sunday, etc.). It is likely that pricing depends 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.

For reasons of logical consistency, we model the random effects of marketing mix variables with their expected sign, by using a log normal distribution. For example in the case of price effects, β_i is modeled as,

$$-\beta_i = \exp\left(\beta_0 + \sigma_\nu' v_i\right), v_i \sim N\left(0, 1\right) \tag{2}$$

In practice, consumers can choose among several choice options, the so-called "inside goods," or

tion 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.

[8] In contrast, while the random intercept model can be specified with a full variance-covariance matrix, in practice it often it is constrained to be diagonal, especially with the use of aggregate data (an exception is Chintagunta, Dubé, and Singh 2002).

decide to buy something else (including "nothing") in a given week, the so-called "outside good." Its utility, u_{i0t}, is fixed at zero 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:

$$\Pr\left(j_t \mid x_{jt}, i\right) = \frac{\exp\left(\alpha_{ij} + x_{jt}\beta_i + \xi_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(\alpha_{ik} + x_{kt}\beta_i + \xi_{kt}\right)} \tag{3}$$

As stated above, our market share time series, purchase set size distribution, and brand penetration data are all different functions or manifestations of these choice probabilities.

To illustrate how they depend on the demand primitives, market level shares s_{jt} are equal to the expectation of the choice probabilities taken over the distribution of individuals i:

$$\begin{aligned} s_{jt} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pr\left(j_t \mid x_{jt}, i\right) \phi\left(v\right) \phi\left(\omega\right) \partial v \partial \omega \\ i &= 1, ..., N, \qquad j = 0, ..., J, \qquad t = 1, ..., T \end{aligned} \tag{4}$$

where $\phi\left(.\right)$ is the PDF of the normal distribution.

The other quantities on which we have data can also be expressed in terms of the same demand primitives and choice probabilities. For example, in the case of brand penetration, the individual level probability that brand j is chosen at least once within a T period time frame is equal to 1 minus the joint probability that the brand is not chosen for T periods. Brand penetration is the expectation of this quantity across individuals,

$$\begin{aligned} \pi_j &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} 1 - \prod_{t=1}^{T} \left(1 - \Pr\left(j_t \mid x_{jt}, i\right)\right) \phi\left(v\right) \phi\left(\omega\right) \partial v \partial \omega, \\ i &= 1, ..., N, \qquad j = 0, ..., J, \qquad t = 1, ..., T \end{aligned} \tag{5}$$

etc.

With this set up, the parameters to be estimated are $\theta = [\alpha_j, \beta_0, L, \sigma]$. The integrals shown in equations 4 and 5 lack an analytical form but can be approximated using simulation methods, described in more detail in the estimation section.

4 Data

4.1 The Frozen Pizza industry

Our empirical analysis covers the Frozen Pizza category. 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

	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
Fringe	0.33	0.31	0.31	0.30	0.31

Table 1: Evolution of average shares for the main brands in the frozen pizza category in each of the years in the data set.

almost 20% of the total pizza business, with delivery pizza being its main competitor outside of the category (*Pizza Marketing Quarterly*). During 1993-1995, the years preceding our analysis, the category was viewed as largely stable, 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).

Two companies, Kraft and Schwan's Food Company, have a dominant place in the Frozen Pizza category and each compete with multiple brands. Kraft's brands include DiGiorno, Tombstone and Jack's while Schwan's 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 volume of the category.[9] All of them, except Jack's, are available nationally. Jack's distribution is limited to markets in the North-West and Mid-West region of the country, but the brand has a large share in those markets.

To avoid 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 1. Nationally, the dynamics in DiGiorno's share reflect a roll-out that took three years. By 1999, DiGiorno had captured about 13% of the U.S. frozen pizza market.[10]

One of the main features of the local market share data is that the launch of DiGiorno apparently

[9] Freschetta is not included in our analysis of the introduction of DiGiorno, because it was introduced later.
[10] The same figure is independently reported in Holcomb (2000).

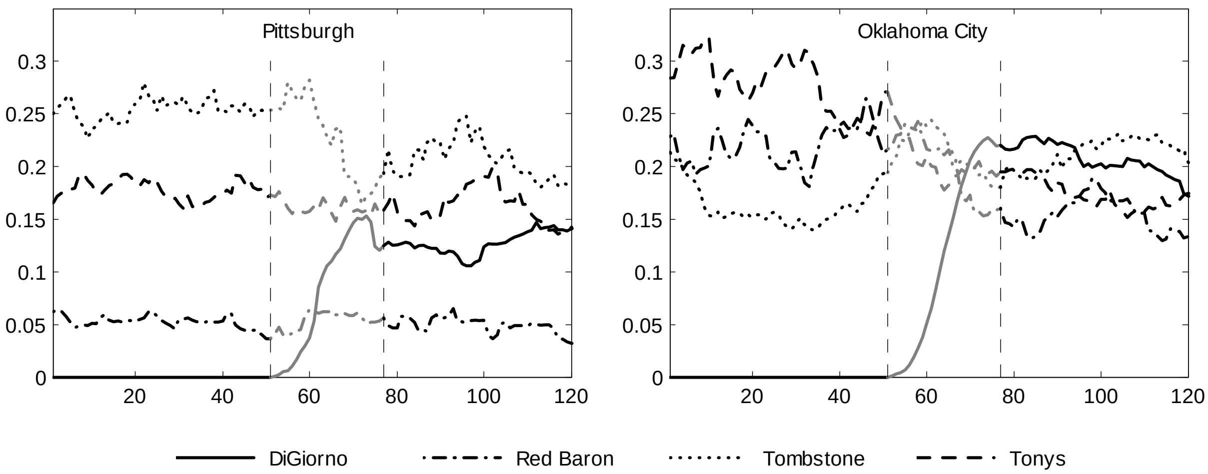

Figure 2: Evolution of market shares of the main brands in two markets at the time of DiGiorno's introduction

affected the incumbent brands differently in different markets. In Figure 2, we show the evolution of market shares in Pittsburgh and Oklahoma City, around the time of DiGiorno's introduction. In Pittsburgh, DiGiorno steals most share from Tombstone whereas the share of Red Baron (and that of Tony's) is not at all affected by the launch of DiGiorno. Oklahoma City presents almost the opposite situation. Red Baron (and Tony's) suffers a severe share loss, but Tombstone's share remains unaffected.

These differences suggest that substitution patterns may be market specific. One expectation about local substitution patterns is that differences across markets arise because the incumbents' shares are different (see, e.g., Bronnenberg, Dhar, and Dubé, 2007) and new brands draw their share proportionally from those incumbents. But this proportional draw hypothesis does not accord well with the data. For instance, in Oklahoma City, the rank-order of brand size changes for the incumbents, thereby dispelling substitution patterns consistent with proportional draw. Indeed, share erosion for the incumbent brands is concentrated in the Red Baron and Tony's brands, and Tombstone, which was smaller than these brands prior to launch of DiGiorno, actually becomes larger after it. These patterns are suggestive of more complex local substitution patterns. Our model is capable of capturing these through random tastes with a correlation structure.

4.2 Alternative data sources

Our analysis integrates three different data sets. The first data set covers market-level sales volume, price, local feature advertising and display levels. The data cover 260 weeks, from January 1995 to December 1999, and 64 IRI markets. Markets are defined by IRI as metropolitan areas or as part of a state. The sample of 64 markets covers the lower 48 states of the US. The data are the result of aggregation over a sample of stores in each 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 post launch sales that often reflect 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 post-launch market shares given distribution. Thus, for the empirical analysis, we censor the 26 week period after launch of DiGiorno (see Figure 2). This means that every market is represented by two time series, one representing the situation before and another the situation after the launch of DiGiorno for a period of 52 weeks.

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

The third set of data consists of summary statistics of purchase behavior, compiled by the AC Nielsen company, using its HomeScan panel. From these data, we have access to the local distribution of purchase set sizes, i.e., the percentage of consumers that buy 1, 2, 3, 4, or more unique brands over the duration of a year. We have these data for each of the 9 Census divisions[11] and for the year 2004. We also have access to annual category penetration levels at the Census divisional level, for the years 2000 to 2003, measuring the percentage of people that have purchased pizza at least once during each year.[12] Finally, we have the same measures at the brand level for the same years. Each of the 64 markets maps into exactly one Census division.

5 Demand estimation

5.1 Overview

We estimate the demand model using the Generalized Method of Moments (GMM) algorithm (see, e.g., Hansen 1982). The GMM algorithm combines different sets of information assigning optimal weights to each piece of data (Lancaster and Imbens, 1994), while allowing the use of

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

[12] A point of concern is that these data were collected a few years later than the sales data. However, the penetration measures are very stable across time. This limitation is one of the data and not of the approach or of its practical scope. We tested the robustness of our findings against assuming that the penetration and purchase set size data are observed with error and found that our results do not change substantively.

instrumental variables to correct for the correlation that is generally present between price data and unobservable demand shocks (see also Petrin, 2001). The possibility to combine multiple data sets and use instrumental variables makes the estimation method ideal for our purpose.

We use three different sets of moment conditions. First, we use moments similar to Berry, Levinsohn, and Pakes (1995), and Nevo (2001). These moments require that the demand shocks, ξ_{jt}, are orthogonal to the instrumental variables.

Second, in many studies, the size of the outside good is based on an assumption that is more or less ad hoc. In this study, we define moment restrictions combining the weekly dynamics in the local Frozen Pizza sales and in the local category and brand penetration rates to estimate the size of the outside good empirically.

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. Whereas the "BLP moments" are found to be effective at capturing the mean utilities in the market, we find they do not capture heterogeneity well. Not surprisingly, this also affects the inference of the origins of the share for new entrants and trial rates.

We now present the details for the implementation of each of these moments.[13]

5.2 The BLP moments

In empirical studies of demand, the analyst often lacks observation of certain demand primitives that 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 instrumental variables and impose an orthogonality condition with the unobserved demand shocks ξ_{jt}. Berry, Levinsohn, and Pakes (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, $\delta_{jt} = \alpha_j + \xi_{jt}$.[14] Next, given an initial value δ_{jt}^0 and a set

[13] We do not model the supply side, i.e., prices, to help estimate the demand parameters. In order 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 multi-product 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, Siegfried, and Howell 1992).

[14] Recall that the β_i have a log normal distribution. Hence, we can not factor out the population mean as a linear

of parameter values, the following expression is iterated until it converges[15]

$$\delta_{jt}^{n+1} = \delta_{jt}^{n} + \ln\left(s_{jt}\right) - \ln\left(\widehat{s}_{jt}\left(\delta_{jt}^{n}, \theta\right)\right), \tag{6}$$

where $\widehat{s}_{jt}$ is the predicted market share, which integrates over all consumer types, s_{jt} is the actual share, and n counts the iterations in the BLP contraction mapping of equation (6). Once δ_{jt} are computed, we can estimate ξ_{jt} as the residual of the equation $\delta_{jt} = \hat{\alpha}_j + \xi_{jt}$, where in turn $\hat{\alpha}$ can be estimated.

The shares s_{jt} are not actually observed. Instead, what is observed is the category sales volume over time and the share among the inside goods, i.e., the conditional shares $\tilde{s}_{jt} = s_{jt}/\left(1 - s_{0t}\right)$. 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 about s_{0t}. Our implementation replaces s_{jt} in the estimation by the share among the inside goods, $\tilde{s}_{jt}$, which is data, multiplied by 1 minus the estimated share for the outside good (see next subsection).

$$s_{jt} = \tilde{s}_{jt} \times \left(1 - \widehat{s}_{0t}\right) \tag{7}$$

Given our additional moment restrictions below, this suffices for identification.[16]

As instruments, we use prices in other markets (as in Nevo 2001). A concern with such instruments is that they still correlate with the demand shocks if prices are set in clusters of markets rather than in each market individually. To alleviate this concern, we use as instruments prices in far away markets.[17] Related to identification of the positions in the brand map, we impose that the positions L are chosen such that the ξ_{jt} are zero-mean for all brands before and after the launch by DiGiorno. Finally, the instruments for display and feature are the variables themselves. Arraying these instruments into the vector Z_{jt}, we write the "BLP moments" as:

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

where the expectation is taken over products and time.

parameter, to be included in the δ_{jt}, as Nevo (2000) does.

[15] Convergence is obtained $\left|\delta_{jt}^{n+1} - \delta_{jt}^{n}\right| < \varepsilon$, for $\forall \delta_{jt}$, with ε very small, i.e., $10E-09$ in this study.

[16] This modification in and by itself is 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 $\hat{s}_{0t}$ that is contained in it. In our case, we estimate this quantity.

[17] Specifically, for the market of interest m, we sort all other markets in terms of geographic distance to market m. Prices in markets distanced further away than the 10th closest market and on which same period prices are available are potential instruments. We use the prices in 3 far away markets as instruments.

5.3 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, p. 527) 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 strongly 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 Frozen Pizza expenditure (FPE_t) in a market[18] and we then estimate the –non-structural– proportionality or scaling factor using data.

The proportionality factor is determined as follows. In equation (7), we replace $\widehat{s}_{0t}$ by $1 - \lambda \times \mathrm{FPE}_t$, and we define a moment that chooses the scaling factor λ such that the model is consistent with observed brand and category penetration.[19] Brand and category penetration identify λ, because –as in Nevo's observation above– different $\widehat{s}_{0t}$ will generate different parameters θ which in turn imply different penetration rates.[20] Thus in estimation, the structural parameters θ are a function λ, i.e., write $\theta(\lambda)$. Extant papers have the same conditionality and our paper differs from those in that we estimate λ rather than assume it.

To evaluate the moment restriction, we need to compute the category penetration rate implied by the model. This can be done as follows. For each simulated household i (resulting from a draw of v_i and w_i), brand j, and week t, the model predicts a choice probability $\mathrm{Pr}_{ijt}(\theta)$. Write the purchase set as $\{\cdot\}$. This set contains all brands that were bought by the consumer at least once during a year. The probability that a consumer chooses only the outside good over the year after launch of DiGiorno is

$$\mathrm{Pr}_i(\{\varnothing\}) = \prod_{t=T-51}^{T} \mathrm{Pr}_{i0t}(\theta, \lambda, X_t). \tag{9}$$

This compound probability is smooth in θ and in λ. We require that the population mean of this

[18] Thus, category volume is measured as the share of Frozen Pizza expenditure (FPE_t) among all categories scanned in a given market. Alternatively, one can make the joint share of the inside goods proportional to category sales. However, total recorded category sales in our data is 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.

[19] 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 non-structural scaling constant that translates data about category size into "data" about the outside good.

[20] A potential concern is uniqueness because "few" disloyal consumers may lead to the same brand penetration as "more" loyal consumers. However, this distinction is identified by the category penetration condition and by the heterogeneity moments to be explained.

probability, $E\left[\Pr_i\left(\{\varnothing\}\right)\right]$, is equal to 1 minus the actual category penetration rate, π_c. Similar equations can be formulated (see also equation 5) for the model's predictions about brand penetration, i.e., the population mean of the probability that the purchase set contains j, $E\left[\Pr_i\left(j \in \{\cdot\}\right)\right]$. Our moment restriction is that we match these predictions to actual brand penetration π_j. Arraying these $J+1$ conditions, we write the "outside good" moment as

$$G_2\left(\theta, \lambda\right) : E \begin{bmatrix} \Pr_i\left(\{\varnothing\}\right) \\ \Pr_i\left(1 \in \{\cdot\}\right) \\ \vdots \\ \Pr_i\left(J \in \{\cdot\}\right) \end{bmatrix} = \begin{bmatrix} 1-\pi_c \\ \pi_1 \\ \vdots \\ \pi_J \end{bmatrix}, \tag{10}$$

where the expectation is taken over individuals i.

5.4 The heterogeneity moments

We use our data on the distribution of purchase set sizes, S_i, to help 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, $\Pr\left(S_i = 0\right)$, $\Pr\left(S_i = 1\right)$, $\Pr\left(S_i = 2\right)$, etc., in different regions in the United States. For example, in the Pacific Census division, 39% of households buy 0 Frozen Pizza brands in a year (therefore category penetration is 61%), 22% of households buy only 1 unique brand, 17% switch among 2 brands, 10% switch among 3 brands, and 5% switch among 4 brands (the remaining 6% of households switch among more than 4 brands).

We recursively compute the predicted purchase set size distribution of the model from the implied choice probabilities, $\Pr_{ijt}\left(\theta, X_t\right)$ in equation (3). As an example, we provide details on the model's predictions for $\Pr\left(S_i = 1\right)$ and $\Pr\left(S_i = 2\right)$.

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

$$H_{ij} = \prod_{t=T-51}^{T} \left[\Pr_{ijt}\left(\theta, X_t\right) + \Pr_{i0t}\left(\theta, X_t\right)\right]. \tag{11}$$

This probability covers all purchase histories that combine any positive number of purchases of j with any positive number of purchases of the outside good. Therefore (using the notation in equation 9),

$$\Pr_i\left(\{j\}\right) = H_{ij} - \Pr_i\left(\{\varnothing\}\right) \tag{12}$$

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 1, is equal to the summation of $\Pr_i\left(\{j\}\right)$ across choice options $j \neq 0$.

$$\Pr(S_i = 1) = \sum_{j=1}^{J} \Pr_i(\{j\}). \tag{13}$$

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

$$H_{ijk} = \prod_{t=T-51}^{T} \left[\Pr_{ijt}(\theta, X_t) + \Pr_{ikt}(\theta, X_t) + \Pr_{i0t}(\theta, X_t)\right]. \tag{14}$$

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 than the outside good is then (using equation 12),

$$\Pr_i(\{j,k\}) = H_{ijk} - \Pr_i(\{j\}) - \Pr_i(\{k\}) - \Pr_i(\{\varnothing\}). \tag{15}$$

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 $k \neq 0$.

$$\Pr(S_i = 2) = \sum_{j=1}^{J} \sum_{k=j+1}^{J} \Pr_i(C_{iT} = \{j,k\}) \tag{16}$$

The probabilities $\Pr(S_i = 3), \Pr(S_i = 4)$, etc., are recursively computed in a similar fashion.

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

$$G_3(\theta) : E\left[\Pr(S_i = s)\right] = \mathcal{F}_s,\ \ s = \{1, ..., 4\}, \tag{17}$$

where the expectation is again taken over households.[21]

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.

5.5 Objective function and simulation

The objective function combines the three sets of moments previously described:

$$G(\theta) = \begin{bmatrix} G_1(\theta) \\ G_2(\theta) \\ G_3(\theta) \end{bmatrix}. \tag{18}$$

[21] The purchase set size is only matched on a 52 week period of the post-entry data, avoiding the disadvantage of an assumption that the purchase set size remains constant with the introduction of a new product.

In order 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 (17)

$$E[\Pr(S_i = s)] = \int\int \Pr(S_i = s)\,\phi(\upsilon)\,\phi(\omega)\,\partial\upsilon\partial\omega \tag{19}$$

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

$$E_i[\Pr(S_i = s)] \approx \frac{1}{N}\sum_{i=1}^{N} \Pr(S_i = s|\theta, X_t, \upsilon_i, \omega_i) \tag{20}$$

This approximation is again smooth in the parameters θ.

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

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\left(\widehat{G}(\theta)' W\left(\widetilde{\theta}\right)' W\left(\widetilde{\theta}\right)\widehat{G}(\theta)\right) \tag{21}$$

where $\widehat{G}(\theta)$ is the sample analogue of $G(\theta)$ and $W\left(\widetilde{\theta}\right)$ 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 $\widetilde{\theta}$, a preliminary consistent estimate of θ.

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

$$W_1\left(\widetilde{\theta}\right)' W_1\left(\widetilde{\theta}\right) = \left[\frac{1}{T^2}\sum_{t=1}^{T} g_{1t}\left(\widetilde{\theta}\right)\cdot g_{1t}\left(\widetilde{\theta}\right)'\right]^{-1} \tag{22}$$

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

Under the assumptions of the model, the variability of 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 from 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 $\xi_{jt}\left(\widetilde{\theta}\right)$ and compute the empirical value of $G_2\left(\tilde{\theta}\right)$, and $G_3\left(\tilde{\theta}\right)$. 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.[22] The inverse of this matrix is the desired weight matrix $W_2\left(\widetilde{\theta}\right)' W_2\left(\widetilde{\theta}\right)$.

Finally, the complete weigh matrix $W\left(\tilde{\theta}\right)' W\left(\tilde{\theta}\right)$ is the block-diagonal combination of the two parts defined above.

[22] 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 neglibible and can be made arbitrarily small by increasing the number of simulation draws. We tested alternative measures of variance, with similar results.

5.6 Computing local switching

As stated in section 2, we compute local switching after estimating the demand system in each market separately. In order 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. 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:

$$\begin{aligned} \Delta s_{jt} &= \frac{1}{N}\sum_{i=1}^{N}\left[\Pr_{ijt}(\theta, X_t, \text{DiGiorno in}) - \Pr_{ijt}(\theta, X_t, \text{DiGiorno out})\right] \qquad (23) \\ j &= 1, .., J, j \neq \text{DiGiorno} \qquad \forall t \text{ after DiGiorno's entry} \end{aligned}$$

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.[23]

6 Monte Carlo simulation

6.1 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. We generate data according to the utility model (1) and probability model (3) in the paper. This creates an $N \times B \times T$ table of choice probabilities for N simulated households, B brands, and T time periods. For the generation of the data, we choose $N = 10,000$ households, $B = 6$ brands, $T = 104$ weeks. During the first 52 weeks 5 brands are present. A single new brand is launched in week 53. The actual prices and promotion data from 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 brand effects are brand specific. The values for the data generating parameters of the demand model are set at values similar to the empirical estimates for Chicago.

To approximate the demand integrals, we use 500 pseudo households.[24] Starting values for the

[23] An other option is to compare the market shares of the incumbent brands pre- and post-launch 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) and 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.

[24] In addition to using 500 draws in the simulation, we ran models with 250, 500, and even 1000 draws, with no apparent difference in estimation results. In the empirical section of the paper, we therefore use 250 draws to approximate the demand integrals.

Experimental cell	1	2	3
Information	augmented		standard
Share of inside good	known	estimated	known
MAD temporal	0.0243	0.0243	0.0245
MAD purchase set	0.0187	0.0191	0.0654
MAD brand penetration	0.0250	0.0260	0.0805

Table 2: Fit measures in the data experiment

GMM algorithm are computed using a non-linear least squares 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 non-linear least squares estimator converges fast but does not account for price endogeneity. It is therefore only used to obtain starting values for the GMM estimator.

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 cell, 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 cell, 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.

The generated data, the household draws, and the starting values of the demand parameters are kept constant across each experimental "condition." This facilitates comparison of the results across cells.

6.2 Results

We ran the experiment 50 times and saved the point estimates of the demand system for discussion. We first comment on model fit. Table 2 shows that all experimental conditions have the same mean absolute deviation (MAD) in market share fit. That is, 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 cells 1 and 2, we observe a good fit between the model and the data on purchase set sizes and brand penetration,[25] the fit is poor in cell 3, where we only use the time-series market

[25] For example, actual brand penetration is in the range of .07 and .44 across brands. Hence, a mean absolute deviation of .025 between the model's predictions and the data is small compared to the range of brand penetration.

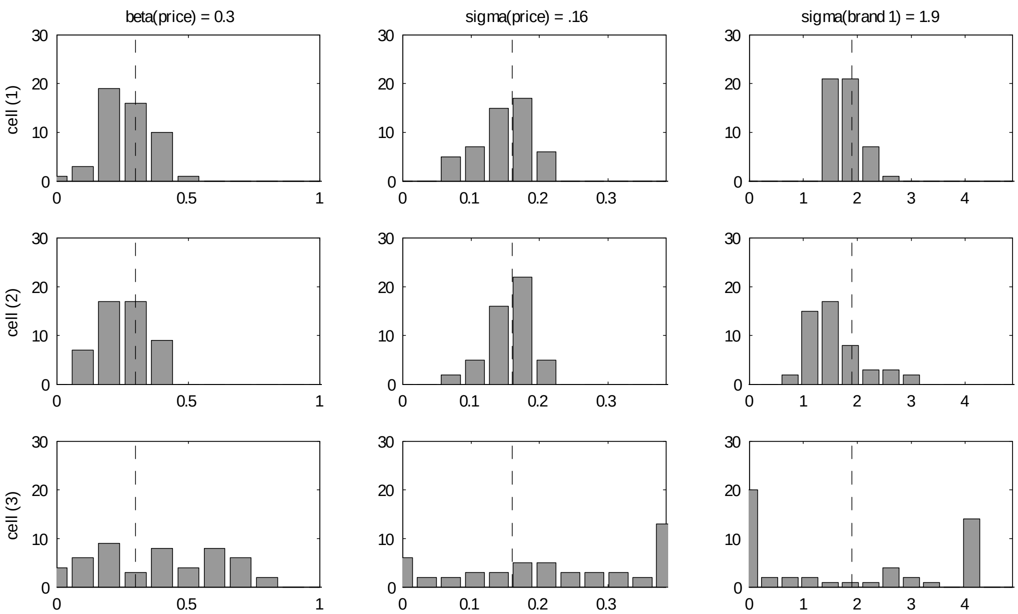

Figure 3: Three key parameters across the experimental conditions.

share data to estimate the model parameters. It is perhaps not unexpected that the augmented information leads to better fit of the purchase set size and brand penetration data. But, what is surprising is that poor fit in terms of purchase set sizes, and brand penetration – as evidenced in cell 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 is 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 3 shows the histograms of the point estimates for the location parameter (β_0) in the lognormal distribution of price coefficients (data generating value is 0.30),[26] heterogeneity in price responses (0.16), and variance of the random effects for brand 1 (1.90). First, we can conclude that the heterogeneity parameters displayed are well identified using the augmented information (cells 1 and 2). The same is true for the variances of the remaining five brands on the market, with data generating values ranging from 1.1 to 2.4. Focusing on cell 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.

[26] 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)

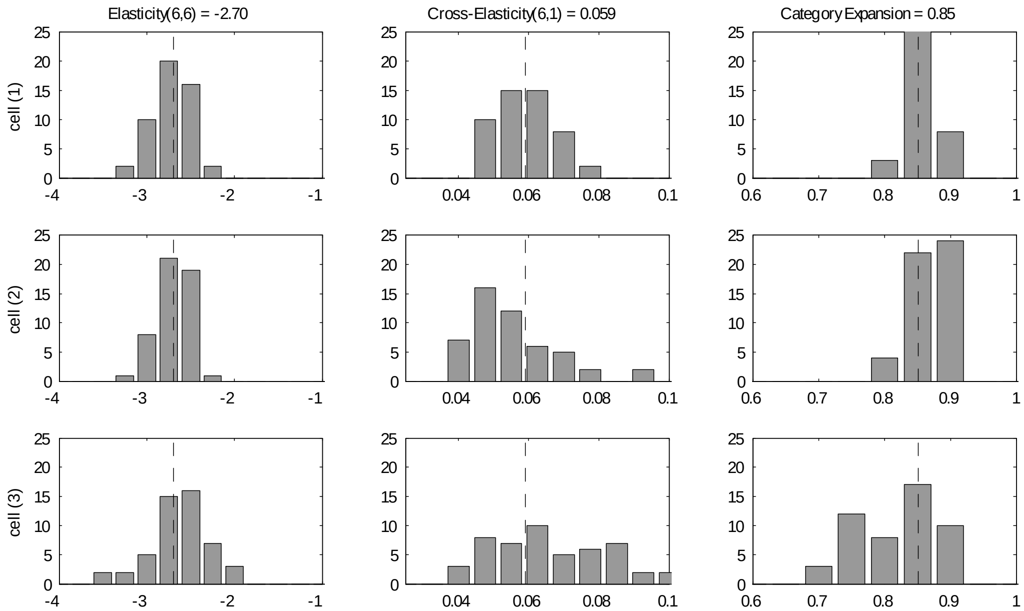

Figure 4: Demand characterstics across the experimental conditions

Contrasting these observations with cell 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 to an estimation upper bound that was set for practical reasons in the experiment (0.4 for price heterogeneity, 4 for brand heterogeneity). These bounds are not binding in cells 1 and 2.

To see how differences in the demand parameters translate in differences in demand characteristics, we post-processed the 50 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 4 shows the histograms of the point estimates of these quantities. It also shows the average of the actual values of the quantities across replications (again with a hatched line). The results in cell 1 show that (cross) elasticities and category expansion are all centered around their actual values. When we estimate the outside good (cell 2) rather than assuming it, the variance of the cross-elasticity estimate becomes higher, but the estimates for own elasticity are virtually identical. Also, the implied fraction of demand that is drawn from the outside good is very similar. Thus, we conclude that the share and the augmented information collectively is useful in identifying (cross) elasticity and category expansion.

However, in cell 3, where we use only market share data, estimates of both the cross as well as own elasticities have much more variance across replications and the cross-elasticity is too high. In addition, the estimates of category expansion are inconsistent and too small when ignoring the purchase set size and penetration data. The latter is due in large part to the overestimation of price heterogeneity in many cases (see Figure 3) which creates a large tail of price sensitive "outside" consumers who do not want to buy 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 very poorly identified from the market share data by the orthogonality conditions in GMM.[27]

We note that we have conducted additional simulation studies including (1) using the data generating parameters values to initialize the estimation, (2) using smaller variances for the random demand shocks ξ, and (3) using the same draws for simulation and estimation (such as to eliminate simulation error entirely). In all of these additional simulation studies, the heterogeneity parameters remain unidentified without the added information on purchase set size and brand penetration.[28]

7 Empirical results

7.1 Introduction

We estimate our model using data from the Frozen Pizza category in each of 64 markets and present our results in the following order. First, we report on the fit of the model. Second, we discuss several structural parameter estimates. Third, we discuss the implied origins of demand for the newly launched brand, DiGiorno. Finally, we discuss the cross-market differences in substitution patterns.

7.2 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 penetration. In addition, to evaluate the improve-

[27] Ours is a model 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.

[28] It is perhaps also interesting to note that the non-linear least squares estimator without the additional data performs better in re-estimating the heterogeneity distributions than GMM with the additional data. Thus, if endogeneity is not considered a major problem, it appears that an NLS estimator has better properties than the GMM estimator.

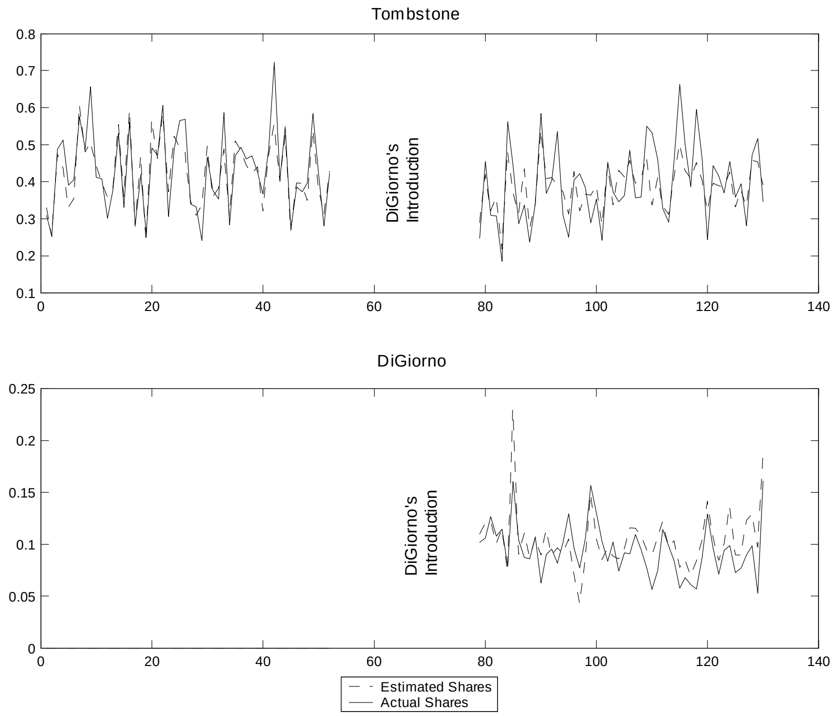

Figure 5: Actual and estimated shares for Tombstone and DiGiorno in Chicago.

ment stemming from the additional moments, we compare our proposed model (the "augmented" model) 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.[29]

To illustrate the findings, we discuss the case of Chicago. Figure 5 displays the time series of actual and estimated shares for 2 brands in this market.[30] It is clear that the model does very well at explaining the brand level variation in market shares. This is also true in the other markets. As in the data experiment, the standard model also does well in recovering the market share time series.

Next, Table 3 shows that the demand primitives obtained from the augmented model imply the purchase set sizes and brand penetration 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. Specifically, the last column of the table shows that the standard model greatly

[29] 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 likely to be conservative.

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

		actual	augmented model	standard model
purchase set size	0 brands	0.238	0.283	0.144
	1 brand	0.214	0.249	0.502
	2 brands	0.170	0.206	0.264
	3 brands	0.132	0.148	0.071
	4 brands	0.096	0.071	0.000
brand penetration rates	Tombstone	0.339	0.369	0.767
	Red Baron	0.359	0.341	0.178
	Tony's	0.161	0.161	0.254
	Totino's	0.145	0.127	0.026
	DiGiorno	0.303	0.282	0.026

Table 3: Implied purchase set size and penetration rates

overestimates single brand loyalty in this market. At the same time it underestimates the fraction of households that are loyal to the outside good and the fraction of households that switch among many products.

The brand penetration data show even stronger differences. The augmented model fits the brand penetration data very well. However, the estimates from the standard model are very inconsistent and highly unlikely. For instance, the standard model implies that Tombstone is bought at least once a year by 77% of the market, whereas DiGiorno has a penetration of only 3% of the market. What drives this unlikely difference is that using the standard model, the brand level random effects are inferred to have a variance of almost 0 in the case of Tombstone, and of 10 (a practical upper bound set by the authors) in the case of DiGiorno. Thus, as in the Monte Carlo study, the underlying problem with the standard model is that the heterogeneity parameters are not well identified. The model implies that the demand for DiGiorno comes from few customers that are very loyal, whereas the demand for Tombstone comes from many customers that buy the brand infrequently. These implications of the standard model are in disagreement with the data. Empirically, the market share data by itself appear to be uninformative under the orthogonality conditions $E\left(\xi_{jt} \otimes Z_{jt} = 0\right)$ about brand penetration and trial (in the case of DiGiorno). This observation is not specific to Chicago and applies equally to the other markets (see also below).

7.3 The impact of the extra information

As in the Monte Carlo study, the estimated degree of consumer heterogeneity is very different between the standard model and the augmented model. One measure of how they are different is the frequency with which the estimated heterogeneity parameters are "unreasonable." In the

switching	augmented model	standard model
cannibalization	0.073	0.030
competitive draw	0.047	0.256
category expansion	0.881	0.714

Table 4: Estimates of the switching to DiGiorno, for the market of Chicago, using the augmented model and the standard model, as a percentage of DiGiorno's share.

standard model, the estimates for brand variances (i.e., the diagonal elements of LL') are 0 at least once in 56% of markets, and increase boundless to an upper estimation limit of 10 at least once in 42% of markets. By the same criterion, price heterogeneity is either estimated to be absent in 39% of markets, or very high in 17% of markets.

In contrast, with the augmented model, across 64 markets and 4-6 brands per market, exactly 1 brand heterogeneity estimate and exactly 1 price heterogeneity estimate was at a lower estimation limit of 0. In no cases, was the upper estimation limit binding.

Under the augmented information, the mean (across markets) variance in consumer tastes for Tombstone is 2.71 (with variation across markets $s^2 = 1.32$). For the model using the standard information, the mean variance in tastes is about the same, 2.55, but the heterogeneity estimates differ widely across markets ($s^2 = 7.95$). Using the augmented information the mean variance in price effects is 0.73, and the estimates of this parameter have a cross market variance of $s^2 = 0.73$. Under the standard model, these numbers are 1.12 and 2.34 respectively, suggesting a higher (and more inconsistent) degree of price heterogeneity than was inferred using the augmented model.

The differences in the estimates of heterogeneity directly impact which brands form close substitutes to the new brand, DiGiorno. Therefore they also impact the estimated origins of DiGiorno's share. This is exemplified in Table 4 for Chicago (we generalize to other markets subsequently). In the Chicago market, the standard model displays a lower percentage of switchers from the outside good to DiGiorno than the proposed augmented model (71% vs 88%). The competing brands are a large source of demand for DiGiorno under the standard model, but not under the augmented model. The augmented model has more face validity given (1) that DiGiorno's aim was to steal demand from the outside good, and (2) that this approach implies the correct number of brands that consumers switch among and the correct brand penetration rates.

Table 5 helps to form an intuition for why the additional information yields better estimates of heterogeneity. It uses data from the Oklahoma City market, again as a representative example, and lists the estimated brand level variances (heterogeneity) for the main brands. It also displays the

	heterogeneity	penetration	share	share/penetration
Tombstone	2.104	0.181	0.242	1.342
Red Baron	0.743	0.408	0.174	0.426
Tony's	2.531	0.186	0.202	1.085
Totino's	1.181	0.261	0.096	0.368
DiGiorno	2.388	0.264	0.231	0.878

Table 5: Purchase behavior and preference heterogeneity

observed penetration rates and market share data, along with the ratio of the share and penetration data. Observe that Tombstone has a high share relative to its penetration, whereas Totino's has a low share relative to its penetration. Thus, relatively speaking, a brand like Tombstone appeals to few customers who like the brand a lot (given its relatively high share), whereas Totino's appeals to many customers. This means that the preferences for Tombstone must be more dispersed in the population than the preferences for Totino's. Indeed, our heterogeneity estimates are consistent with this. The correlation between the heterogeneity estimates and the ratio between share and penetration (second and fifth columns of Table 5) is 0.82. We computed a similar correlation in each market. On average this correlation is 0.725, confirming that there is a very strong correlation between observed share/penetration ratios and estimated heterogeneity.

In conclusion, we believe that the additional information helps create better and more intuitive estimates of brand heterogeneity.

7.4 Structural parameters and brand perceptions

We now report on selected demand parameters and characteristics. Because demand parameters were estimated for each of 64 markets, we summarize the results rather than discuss them for each market separately.

The market for Frozen Pizza is fairly price sensitive. Own elasticities for the new brand have a mean of -1.92 across markets, with a variance $s^2 = 0.73$. We note that the elasticities when estimated using the standard information are markedly lower (-1.46, $s^2 = 1.07$), which is consistent with the higher degree of inferred price heterogeneity.[31] The display and feature dummies have effect sizes of 1.14 ($s^2 = 0.19$) and 0.44 ($s^2 = 0.14$), respectively.

The estimation of the outside good share is fundamental in the process of accurately identifying switching between alternatives, since in many cases category expansion plays an important role in generating demand for a new brand, especially one that offers unique product features. In the case

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

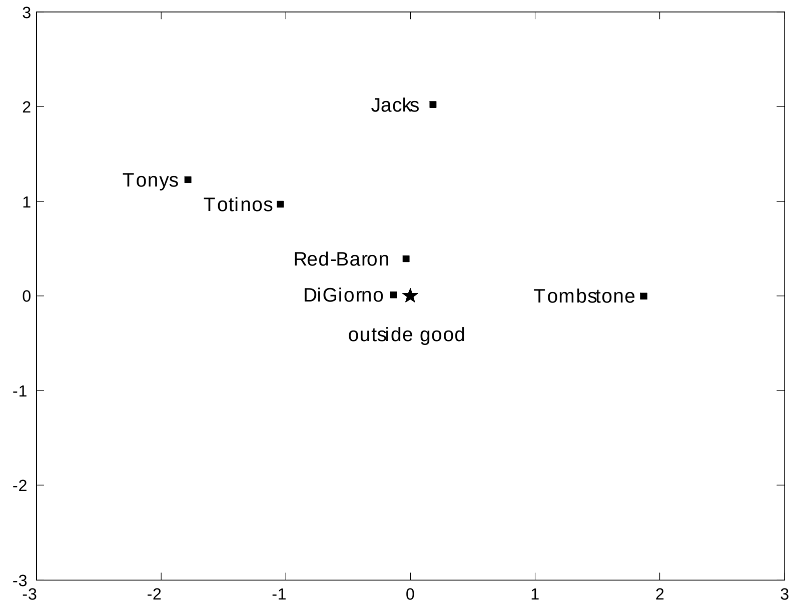

Figure 6: Perceptual maps for the main brands in Chicago

of the frozen pizza category, the outside alternative has additional relevance, since DiGiorno used a well documented media campaign that specifically targeted delivery pizza, a component of the outside good.

Our estimate of the outside good is based on market level penetration rates. Using our approach, we find an estimated weekly share of the outside good that ranges across markets from a minimum of 77% to a maximum of 94%. *Prima facie* these results are plausible and accord with the industry experts' estimate that the Frozen Pizza category constitutes approximately 20% of the total Pizza Category (*Pizza Marketing Quarterly*).

Our approach also produces estimates of the perceptual brand map. We allowed for the existence of two unobserved attributes, leading to a two-dimensional "perceptual" map. Figure 6 shows the estimates for the market of Chicago. A few findings are worth pointing out. First, consumers perceive that DiGiorno is located closest to the outside good in this market. This seems consistent with DiGiorno's positioning.[32] Second, in Chicago, DiGiorno was successful at positioning away from Tombstone and Jacks (Kraft's other brands) which suggest the absence of strong cannibalization (see also Table 4). We note however that brand perception maps are highly market specific.

[32] But, it should be noted that the perception map is reflective of correlations between brand level random effects, net of all observed effects.

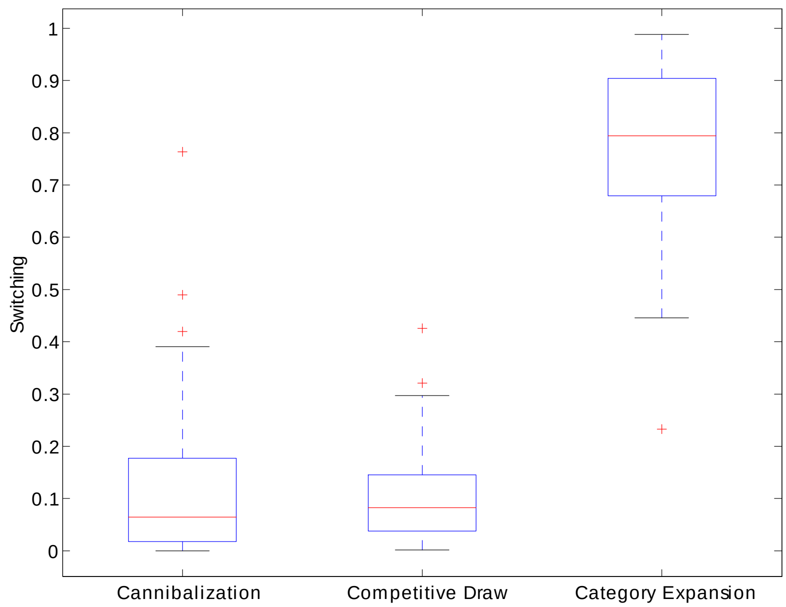

Figure 7: Distribution of the origin of switching across 64 markets

7.5 Comparing switching behavior across markets

In this section, we conduct a policy experiment by removing DiGiorno from each market and estimating the adjustment in market shares for the remaining brands. Figure 7 summarizes the inferred origins of demand for DiGiorno from this experiment across the 64 IRI markets, using the augmented data. From this experiment, the demand for the new brand comes mostly from category expansion. The average fraction of DiGiorno's demand that comes from the outside good is 77.1%. Table 6 shows that when we use the standard approach, this fraction is estimated to be less (63.2%). In contrast, the latter model overestimates the amount of demand that originates from cannibalization and competitive draw. Figure 7 demonstrates that there is ample variation in cannibalization, competitive draw and category expansion across markets

To link these estimates to the motivating example involving Pittsburgh and Oklahoma City (Figure 2), Table 6 gives the estimates for the origins of demand for these two cities. The estimates from our policy experiment are consistent with the motivating graphs. From the graphs, in Pittsburgh, an important source of demand is Tombstone (a Kraft brand), while in Oklahoma City, Red Baron and Tony's (both competitor brands) are an important source of DiGiorno's demand but Tombstone's demand is unaffected. Indeed, the policy experiment finds that cannibalization is strong in Pittsburgh but absent in Oklahoma City. The contrast with the standard model is pronounced in the case of Pittsburgh, but less so in the case of Oklahoma City.

		augmented model	standard model
Average across markets	Cannibalization	0.123	0.201
	Competitive Draw	0.106	0.167
	Category Expansion	0.771	0.632
Pittsburgh	Cannibalization	0.390	0.415
	Competitive Draw	0.071	0.218
	Category Expansion	0.539	0.367
Oklahoma City	Cannibalization	0.010	0.029
	Competitive Draw	0.297	0.287
	Category Expansion	0.693	0.684

Table 6: Origins of switching in selected markets

From the results, we conclude that substitution patterns have considerable variation across markets, which is surprising given that the same products are being sold across markets under nationally known brand names. We next visualize several of the geographic properties of this variability in substitution patterns. Figure 8 represents the market-level demand decomposition for the augmented model, whereas Figure 9 represents the same information for the standard model.

Focusing on the first case, Figure 8 shows that demand decomposition is clustered in regions, spanning multiple markets. There are regions of relatively high cannibalization (the North-East and Florida), of high competitive draw (the South-Atlantic states), and of high category expansion (most of the Western United States). New York City is estimated to have the highest degree of cannibalization. Indeed, the pre-launch data show that Tombstone has almost 80% market share, whereas in the past-launch data DiGiorno has 40% share and Tombstone has 40% share. Thus, it appears that in this market, Kraft achieved DiGiorno's share at the expense of another of its brands, Tombstone.

From this graph, we infer that demand composition for DiGiorno is region specific.[33] An important component of brand substitution does therefore not appear to be based on brand names or product characteristics –both traditional drivers of brand substitution– which are constant across markets.

Finally, we observe that both cannibalization and competitive-draw gain in importance as sources of new product demand when we base our inferences on the time series only. The cause of this –in our view– wrong inference is the increased degree of price heterogeneity that is inferred in absence of purchase set size and penetration data. In turn, this produces the inference that the

[33] This is not only true for substitution patterns. Bronnenberg, Dhar and Dubé (2007) report the same is true for CPG brand shares and local dominance in sales.

Figure 8: The geographic distribution of the origins of brand switching using the augmented model

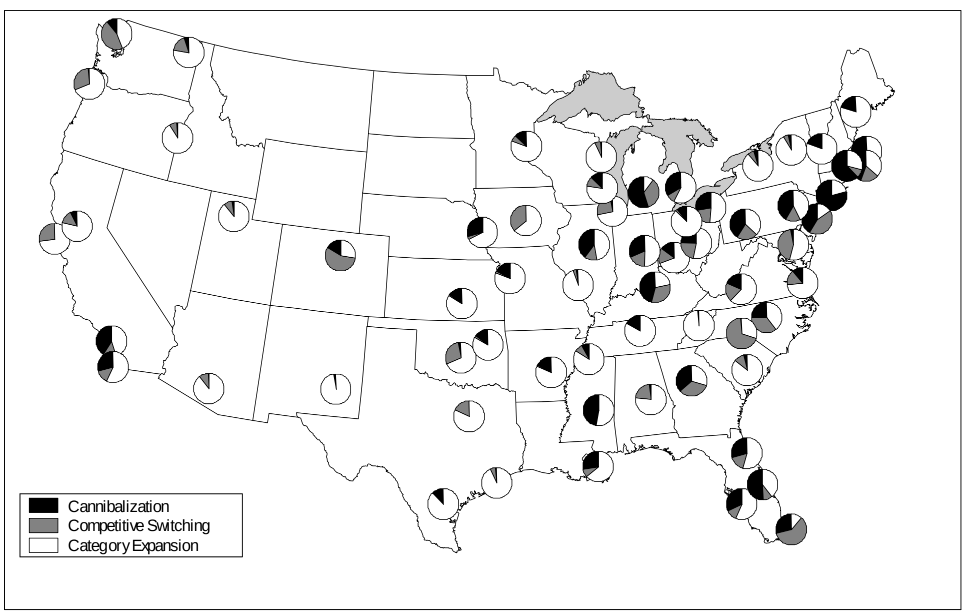

Figure 9: Geographic distribution of the origins of brand switching using the standard model

consumers in the untapped part of the market are very price sensitive. Consequently, the model concludes that there are relatively few consumers that can be attracted to the category with a premium priced brand, all else equal.

8 Conclusion

In this study, we provided an in-depth analysis of local demand for a new national CPG product. Methodologically, we aimed to add insight into resolving a recurring dilemma in the estimation of demand. On one hand, large samples of individuals for all markets/stores under analysis, contain very rich data but are almost impossible 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 on sales or market share, which is a marginal of the individual level data across households, with summary statistics of consumer purchase behavior, which is a marginal of the individual level data across time periods. Technically, the paper offers an improvement to the estimation of heterogeneity within an instrumental variables GMM estimation framework.

Using a Monte Carlo experiment, we have shown that a combination of aggregate level data produces good estimates of individual level preference dispersion in a GMM estimation approach. In contrast, using time series data of market shares gives very 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 demand model is helpful at 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 highly successful CPG brand in the Frozen Pizza category. From company interviews and from its well known media campaign, we know Kraft was focusing its attention 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. However, there are large differences across markets in how successful DiGiorno was at stealing share from the outside good. Similar to what we observe in the outside good, switching from the inside goods is also very different across markets. The cross-market variation in the origins of the demand for the new brand demand implies that substitution among brands is surprisingly not a "national" phenomenon.

Finally, in practical terms, we believe 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 pre-entry stage, based on post-entry data from nearby markets.

We note several avenues for future research. Not much is known at present about why brand substitution is different across regions. For instance, we do not know whether the primitives of brand similarity/substitution are consumer related, retailer related, or related to the marketing actions of the manufacturers. It would be useful to direct a future study in this context. Another avenue for future research is to investigate the broader role of advertising content in shaping the substitution patterns during the new product launch.

References

[1] AC Nielsen (2001), *Consumer Insight: New Product Introductions*, Consumer and Market Trends Report, AC Nielsen, Chicago, IL

[2] Berry, S.T. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25:2, pages 242-262

[3] Berry, S.T. , J. Levinsohn and A. Pakes (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63:4, 841-890

[4] Berry, S.T., J. Levinsohn and A. Pakes (2004). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112(1), 68-105

[5] Blattberg, R. and K. Wisniewski (1989), "Priced-Induced Patterns of Competition," *Marketing Science*, 8:4, 291-309

[6] Bodapati, A. and S. Gupta (2004). "The Recoverability of Segmentation Structure from Store-Level Aggregate Data," *Journal of Marketing Research*, 41:3, 351-364

[7] Bronnenberg, B.J., S.K. Dhar, and J.P. Dubé (2007), "Consumer Packaged Goods in the United States: National Brands, Local Branding," *Journal of Marketing Research*, 44 (February), 4-13

[8] Bronnenberg, B.J. and C.F. Mela (2004), "Market Roll-out and Retailer Adoption for New Brands," *Marketing Science*, 23:4, 500-518

[9] Brownstone, D. and K. Train (1999), "Forecasting New Product Penetration With Flexible Substitution Patterns," *Journal of Econometrics*, 89, 109-129

[10] Buzzell, R.D. and B.T. Gale (1987), *The PIMS Principles*, The Free Press,

[11] Carpenter, G.S., L. Cooper, D.M. Hanssens, and D. Midgley (1988), "Modeling asymmetric competition," *Marketing Science*, 7:4, 393-412

[12] Carpenter, G.S. and Lehmann, D.R. (1985), "A Model of Marketing Mix, Brand Switching, and Competition," *Journal of Marketing Research*, 22:3, 318-329

[13] Chintagunta, P.K. (2001), "Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data,"*Marketing Science*, 20:4, 442-56

[14] Chintagunta, P.K., Dube, J. P. and V. Singh (2002). "Market Structure Across Stores: An Application of a Random Coefficients Logit Model with Store Level Data," *Econometric Models in Marketing,* 16, 191-221

[15] Elrod, T. (1988), "Choice Map: Inferring a Product Market Map From Panel Data," *Marketing Science*, 7, 21-40

[16] Elrod, T., and M. P. Keane (1995), "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, 1-16

[17] Erdem, T. (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15:4, 359-378

[18] Dekimpe, M.G., D.M. Hanssens (1999), "Sustained spending and persistent response: A new look at long-term marketing profitability," *Journal of Marketing Research*, 36(4), 397-412

[19] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50:4, 1029-1054

[20] Holcomb, R. (2000), *Food Industry Overview: Frozen Pizza,* Report FAPC-103, Food and Agricultural Products Center, Oklahoma State University

[21] Imbens, G.W. and T. Lancaster (1994), "Combining Micro and Macro Data in Microeconometric Models," *Review of Economic Studies*, Blackwell Publishing, 61:4, 655-80

[22] Kahn, B. E and L. McAllister (1997), *The Grocery Revolution*, Reading: Mass Addison Wesley

[23] Luo, Lan, P.K. Kannan, and Brian T. Ratchford (2007), "New Product Development Under Channel Acceptance," *Marketing Science*, 26:2, 149-163

[24] MacLeod, W.C. (1995), *Statement of Grocery Manufacturers of America, Inc.*, FTC Hearings on Global and Innovation-Based Competition, http://www.ftc.gov/opp/global/macleod.htm

[25] Nelson, P., J.J. Siegfried, and J. Howell (1992), "A Simultaneous Equations Model of Coffee Brand Pricing and Advertising," *The Review of Economics and Statistics*, MIT Press, vol. 74 (1), 54-63

[26] Nevo, A., (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9:4, 513–548

[27] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry.", *Econometrica*, 69, March, 2001, 307-342

[28] Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4), 705-29

[29] Pizza Marketing Quarterly, *Rising Crust Pizza, Should You Worry?*, http://www.pmq.com/digiorno.shtml.

[30] Shugan, Steven M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 24 (February), 1-18

[31] Sriram, S., Pradeep K. Chintagunta, Ramya Neelamegham (2006), "Effects of Brand Preference, Product Attributes, and Marketing Mix Variables in Technology Product Markets," *Marketing Science*, 25:5, 440-456

[32] Van Heerde, H., C.F. Mela and P. Manchanda (2004), "The Dynamic Effect of Innovation on Market Structure," *Journal of Marketing Research,* 41:2, 166-183

[33] Van Heerde, H., S. Gupta, and D.R. Wittink (2003), "Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is," *Journal of Marketing Research*, 40:4, 481-491

[34] Van Oest, R. and P.H. Franses (2005), "Which Brands Gain Share from Which Brands? Inference from Store-Level Scanner Data," *Quantitative Marketing and Economics*, 3, 281-304