WORKING
PAPERS

Biases in Demand Analysis Due to Variation in Retail Distribution

Steven Tenn
John M. Yun

WORKING PAPER NO. 287

February 2007

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

Biases in Demand Analysis Due to Variation in Retail Distribution

Abstract

Aggregate demand models typically assume that consumers choose between all available products. Since consumers may be unwilling to search across every store in a given market for a particular item, this assumption is problematic when product assortments vary across stores.

1 Introduction

Retail distribution is a requirement for retail sales. One would therefore expect demand studies to account for product availability in a careful manner. In practice, however, the role of retail distribution is implicitly determined by the level of data aggregation employed. Demand studies based on store-level data typically assume consumers limit their purchases to products available at a particular store. Previous research suggests this is a reasonable approximation for certain types of goods, such as grocery products (Rhee and Bell 2002). This highlights a potential problem in demand analyses that use regionally or nationally aggregated data, since they make the opposite assumption.[1] Such studies assume consumers freely choose between all products available in a given market, even items carried by very few stores. They ignore search and transportation costs that may lead consumers to limit their choice sets to a subset of the available items.

The following example illustrates why this approach is potentially problematic. In 2000, the Federal Trade Commission (FTC) challenged the acquisition of Beech-Nut Nutrition Corporation by H.J. Heinz Company, both manufacturers of baby food. The district court judge noted that "nearly all supermarkets stock only two brands of baby food, not three…Gerber is invariably one of the two."[2] The fraction of stores that carried Gerber, Heinz, and Beech-Nut was approximately 100%, 40%, and 45%, respectively.[3] Standard aggregate demand models would fail to control for this distribution pattern if consumers primarily substitute between products available at the same store. One would expect estimated cross-price elasticities between Heinz and Beech-Nut to be close to zero not because consumers are necessarily unwilling to substitute between them, but because few consumers visit stores where both are available.

[1] Throughout the paper, "aggregate data" refers to when sales from multiple stores are combined.

[2] *FTC v. H. J. Heinz Co.*, 116 F. Supp. 2d (D.D.C. 2000) at 193.

[3] *Id*. at 194.

Many demand studies, however, rely on data aggregated across multiple retail outlets (e.g., all stores in a given city). Examples are Hausman et al. (1994), Hausman and Leonard (2002), Berry et al. (1995), Nevo (2001), and Petrin (2002). These analyses assume that consumer choice sets include all products available in a given market, even if an item is available in only a few stores. The plausibility of this assumption depends on the degree to which stores carry different sets of products. Consumers may be unwilling to search across many stores for a given item, especially if the search area contains hundreds or thousands of stores (which is often the case for regionally or nationally aggregated data).

To summarize, product availability plays a very limited role in demand estimation. Strong assumptions are made regarding the relationship between product availability and consumer choice sets without supporting evidence regarding their plausibility. As illustrated by the baby food example described in the introduction, this approach is problematic when there is significant heterogeneity in product availability across stores. The objective of this paper is to determine whether the degree of limited product availability that typically occurs is sufficient to bias significantly the results of aggregate demand models that incorrectly assume all consumers face the same choice set.

3 Data

We utilize weekly scanner data provided by ACNielsen that covers five grocery categories: frozen novelties, shelf-stable pasta, hot dogs, ice cream, and salad dressing. The data reports sales from fourteen retailer-city combinations for the period December 1998 to June 2001 (132 weeks).[6] For each UPC, the dataset reports dollar and unit sales, and the percentage of stores that carry that item. Recognizing that stores significantly vary by size, ACNielsen weights each store by its annual dollar sales (across all product categories) when calculating the percentage of stores where each product is available. This measure, known as "All Commodity

[6] A confidentiality agreement with ACNielsen prohibits retailer names from being revealed.

Volume" ("ACV"), is the standard metric that brand managers and other practitioners use to quantify a product's retail distribution.

In addition, the dataset reports the ACV of each product line, which is a collection of similar UPCs from the same brand. For example, "Ben & Jerry's ice cream" and "Ben & Jerry's frozen yogurt" are distinct product lines. A product line's ACV reports the percentage of stores that carry at least one UPC from that line. In section 6, we use this measure to examine whether UPCs with limited retail distribution are secondary varieties or package sizes of widely available product lines.

4 The Extent of Retail Distribution

Aggregate demand models consider a product "available" if it is sold in at least one store are widely available. Depending on the category, 62% to 81% of dollar sales come from products with a median retail distribution of at least 90% of stores.

Our findings indicate that while supermarket chains carry a fairly homogeneous set of popular items (i.e., widely available, top-selling items), stores have different assortments of low-selling products. To simplify the analysis, researchers commonly exclude low-selling items when estimating demand (e.g., Chintagunta 2002, Nevo and Hatzitaskos 2005). The fact that low-selling items often have low product availability is obviously immaterial when such products are excluded from the data sample.

However, a third set of products with intermediate distribution does pose a problem for demand estimation. Unlike products with low availability, this set of items constitutes a substantial fraction of dollar sales, 16% to 33% depending on the product category. Moreover, they constitute between 19% and 43% of all available products. Items with intermediate product availability are too numerous and too large a fraction of total sales to be reasonably ignored in demand analyses. However, since such products are not carried by a significant fraction of stores, inclusion of these items in aggregate demand models is similarly problematic (as demonstrated by the baby food example presented in the introduction). While a simple point, the literature has failed to recognize this problem.

5 Monte Carlo Analysis

The previous section demonstrated that many products have limited retail distribution. We now explore the implications of this finding, specifically whether aggregate demand models that ignore product availability give biased estimates. Due to the difficulty of analytically determining the bias from estimating a mis-specified model, we rely upon Monte Carlo analysis.

The data used in the Monte Carlo simulations is generated assuming demand is determined by a standard logit framework. This specification is chosen for two reasons. First, it requires a small number of demand parameters. This is a key consideration when estimating demand for a large set of products since other commonly employed specifications, such as the

standard practice and calculate the average elasticity across the subset of consumers for which a given pair of items is in their choice set.

(5.4)

product's average market share across all stores. These two measures coincide when all stores are identical. When that assumption is violated, however, market shares in the two models can substantially diverge.

The last potential difference between the two models relates to price coefficient $_{p}$. If the true model is the heterogeneous store logit, but one instead employs the representative store logit framework, one might expect to obtain a biased estimate for this parameter (although the direction and magnitude of the bias is hard to predict).

It is difficult to determine analytically the net impact of these factors. To provide some intuition for the Monte Carlo results presented below we consider a special case. Suppose product *j* is available only in market *m*, and this product is carried by a fraction ACV_{jm} of stores in that market. In addition, assume that product *j*'s market share is identical across all stores where it is available. Let $e_{jj}^{bias} \quad \tilde{e}_{jj} / e_{jj} \quad 1$ denote the percent bias of the own-price elasticity estimate from the representative store model when the heterogeneous store logit is the correct framework (where e_{jj} and $\tilde{e}_{jj}$ are defined in equations (5.4) and (5.7), respectively). This term simplifies as follows in our simple example, where ${}_{p}^{bias}$ denotes the percent bias in the price coefficient.

$$(5.8) \qquad e_{jj}^{bias} \quad (1 \quad {}_{p}^{bias})\frac{1 \quad {}_{jm}}{1 \quad \frac{{}_{jm}}{ACV_{jm}}} \quad 1$$

Bias in the own-price elasticity is the product of two factors. The first is the percent bias in price coefficient $_{p}$, which can be in either direction. The second term, $\frac{1 \quad {}_{jm}}{1 \quad \frac{{}_{jm}}{ACV_{jm}}}$, depends on a product's market share in those stores where it is available as well as its market share across all stores. This term causes the own-price elasticity to be where it is available. The term $\frac{1 \quad {}_{jm}}{1 \quad \frac{{}_{jm}}{ACV_{jm}}}$ limits to a value of one as a product's market share grows increasingly small, so that the own-price elasticity bias approximately equals $\ _{p}^{bias}$.

While it is not possible to predict the direction of bias, one would expect each product's own-price elasticity to be similarly biased since $\ _{p}$ is not a product-specific parameter. This special case is empirically relevant when estimating the demand for a large number of products, where often no single product has a large market share. In our data, $\frac{1 \quad {}_{jm}}{1 \quad \frac{{}_{jm}}{ACV_{jm}}}$ has a value extremely close to one for most products and is never larger than 1.03.[10] As such, in our Monte Carlo results one would expect the bias in each product's own-price elasticity to approximately equal $\ _{p}^{bias}$. As discussed below, this is what we find.

We now consider the bias in the cross-price elasticity estimate from the representative store model when the heterogeneous store logit is the correct framework. Let $e_{kj}^{bias} \quad \tilde{e}_{kj} / e_{kj} \quad 1$ denote the percent bias in the cross-price elasticity. This term depends on ACV_{kjm}, the fraction of stores in market *m* that carry both products *k* and *j*. In our simple example, e_{kj}^{bias} simplifies as follows.

$$(5.9) \qquad {}_{kj}^{bias} \quad (1 \quad {}_{p}^{bias})\frac{ACV_{kjm}}{ACV_{km}} \quad 1$$

therefore biased towards zero on average, albeit with significant variation depending on the extent of a product's availability.[11]

We now turn to the Monte Carlo analysis, which allows us to consider situations where the bias from ignoring limited distribution cannot be analytically determined. To calibrate the model, we use the scanner data described in section 3 to estimate the heterogeneous store logit framework and obtain estimates for model parameters and . When doing so we restrict the data to products with either intermediate or wide distribution (i.e., products with median availability of at least 50% of stores). Although many products have low availability, they account for a very small fraction of dollar sales (see Table 1). As discussed earlier, the limited retail availability of such products is less of a concern since researchers often exclude low-selling items when estimating demand. To be conservative we follow this practice and remove such products. If we were to include them, however, our results would show an even larger bias from ignoring limited product availability.

Calibration of the heterogeneous store model requires three steps. First, we choose which stores in a market carry a given set of products. We assume each market is composed of 100 stores, and then randomly assign which produciBe(rke(1994) fraces, adef (d)-5(ters)] 75(ea2200610⁄MCID 7 ⁄1E

These estimates are taken as the "true" parameter values from which we construct simulated data for each Monte Carlo simulation. We recognize that the obtained demand estimates might be biased for a variety of reasons, such as endogeneity bias, omitted variables bias, and the use of a restrictive functional form. Nonetheless, it allows us to calibrate the model in a way that approximates a real-world setting, and from which we can conduct a Monte Carlo analysis that by construction does not suffer from any of these potential biases.[13]

We conduct 2,500 Monte Carlo simulations, each of which is carried out as follows. Using the parameter estimates and control variables (but not the unit sales data) from the heterogeneous store logit model, we simulate a ne

go beyond what is reported in aggregate datasets. Specifically, evaluation of equation (5.2)

An alternative approach is to rely upon a representative store model, but then undertake sensitivity analysis regarding the likely bias from doing so. Consider the example discussed earlier for which equation (5.9) reports , the percent bias of the cross-price elasticity estimate from the representative store logit model when the heterogeneous store logit is the correct framework. The degree

6 Explanations for Limited Distribution

In this section, we explore why so many products have limited retail distribution. We do package sizes or varieties of widely available products, that could explain why many stores do not carry such items.

The dataset includes three types of product characteristics that describe each UPC. The first defines the set of UPCs that constitutes a given product line. For example, “Ben & Jerry’s ice cream” and “Ben & Jerry’s frozen yogurt” are distinct product lines. The other characteristics describe each product’s package size and variety. In the ice cream category, for example, variety is measured by characteristics such as flavor (e.g., vanilla), fat-content (e.g., “low-fat”), and sugar-content (e.g., “no sugar added”). Each UPC within a product line is defined by a unique combination of package size and variety characteristics.

We use these characteristics to compute statistics regarding whether the other members of a UPC’s product line are widely available. As indicated earlier, we employ a data sample composed of products with intermediate distribution (i.e., products with median availability of 50% to 90% of stores). The first row of Table 3 reports the fraction of UPCs that are part of widely available product lines. This is defined as product lines where at least 90% of stores carry a UPC from that line (although each store may carry a different assortment of UPCs). A large percentage of UPCs with intermediate distribution, 64% to 93% depending on the category, are part of widely available product lines. The second row of Table 3 shows that depending on the category, 22% to 66% of UPCs with intermediate distribution are from product lines where at least one UPC is widely available. The large difference between the first two rows of statistics demonstrates that even though stores largely carry the same product lines, they carry a different subset of items. Heterogeneity in product assortment is so extensive that it is often the case that no single UPC from a widely available product line is itself widely available.

The third and fourth rows of Table 3 examine whether product assortment heterogeneity is due to differences in product variety or package size. We find that only a small fraction of items are alternative package sizes of widely available product varieties. Depending on the category, this is the case for only 1% to 12% of UPCs. In contrast, 20% to 64% of UPCs are from product lines that contain a widely available item with the same package size, but a different variety. Thus, product assortment differences are primarily due to certain stores not carrying all of the varieties contained within a product line, rather than from stores not carrying secondary package sizes.

New Product Introductions

Another reason many products have limited retail distribution is that the process of introducing and discontinuing items often takes place over many weeks. Even if a product eventually becomes widely available, during a transition period it is typically carried by only a subset of a retailer's stores. To show this, we look at how long it takes a new product to become widely available. One slight complication is that we have only 132 weeks of data, which leads to a truncation problem for products introduced towards the end of the dataset. To avoid truncation bias, we estimate a duration model and then calculate the probability of a given introduction spell length using the estimated model parameters.

We restrict the data sample to newly introduced products, and then estimate the following model. Let A_{jrt} denote an indicator variable for whether product *j* is available in at least one store in retailer-city *r* in week *t* (i.e., $ACV_{jrt} > 0$). Similarly, let W_{jrt} denote an indicator variable for whether product *j* is widely available (i.e., $ACV_{jrt} \geq 90\%$). We rely on a two equation discrete-time duration model where A_{jrt} and W_{jrt} are the dependent variables, respectively, X_{jrt} is a set of observed product characteristics, and Λ denotes the logistic cumulative distribution function.

(6.1) $P(A_{jrt} = 1 | A_{jr,t-1} = 1) = \Lambda(X_{jrt}\beta_1)$

To accommodate either positive or negative duration dependence, the set of control variables X_{jrt} includes a fourth order polynomial in the number of weeks since a product was introduced. In addition, it consists of a set of dummy variables for the calendar month, and a set of retailer-city dummy variables that control for heterogeneity with respect to how quickly retailers introduce or discontinue items. Lastly, we include the four variables shown in Table 3, which control for whether an item is from a widely available product line, whether that line includes a widely available item, and whether that line contains a widely available UPC of the same variety or package size. We construct these measures using data from the week prior to when an item is first introduced. They are included in the set of control variables since the availability of related products potentially speaks to whether a newly introduced item will itself become widely available.

The estimation results from the two logit models are presented in Table 4. It is difficult to tell from the parameter estimates what the fourth order polynomial in the number of weeks since a product was introduced looks like. Therefore, we briefly describe its profile, which is similar across the five product categories. In the first model where the dependent variable is A_{jrt}, the probability of being discontinued is highest for products that have just been introduced. It gradually decreases until approximately 20 weeks following the product introduction, after which the probability starts to increase again. Similarly, when W_{jrt} is the dependent variable, the probability of becoming widely available in a given period peaks at around 10 weeks and then gradually declines. As shown in the table, mixed results are obtained for the remaining control variables that measure the availability of the other items in the same product line. In some categories a new product from a widely available product line is more likely to become widely available itself, while the opposite is true in other categories.

For each product *j*, the set of control variables X_{jrt} and parameters $_1$ and $_2$ are used to predict the likelihood of a given spell outcome and duration. The first panel in Table 5 reports the probability of becoming either widely available or being discontinued in the first year, as well as the probability that the introduction spell has not been completed by the end of the first

While creating an exhaustive list of reasons for why retail distribution changes over time is beyond the scope of this paper, it is worth pointing out that nearly all products undergo extensive intertemporal variation in product availability. To show this we estimate the following model, separately for each product *j*.

(6.2) $\ln(ACV_{jrt}) \quad {}_{jt} \quad {}_{jr} \quad {}_{1jr}t \quad {}_{2jr}t^2 \quad {}_{jrt}$, where $Var(\ {}_{jrt}) \quad {}^{2}_{j}$.

Each product's log distribution (ACV) is regressed against a set of fixed effects for time *t* and retailer-city *r*, and a retailer-city specific quadratic time trend. The root mean squared error (RMSE) from the regression, $\hat{\ }_j$, measures the extent distribution is changing over time after accounting for these factors. We include these controls to demonstrate that distributional changes are idiosyncratic, and are not implicitly controlled for in models that account for underlying trends or seasonality. Since time fixed effects and quadratic time trends are commonly used, that is the specification we chose to employ.

Table 7 presents the distribution of the RMSE estimates obtained from equation (6.2).[16] The results show substantial variation in product distribution over time, with a median RMSE of 9% to 16%, depending on the category. Most products have either intermediate variability (RMSE between 10% and 20%) or high variability (RMSE of 20% or more). Not only is limited product availability extremely common, but the extent of most products' retail distribution significantly varies over time.

7 Conclusion

Aggregate demand models typically assume all consumers in a given market shop at the same "representative store," and therefore face the same choice set. We find little empirical support for this assumption. Across the five grocery categories in our dataset, products with

[16] Equation (6.2) requires a sufficient number of observations to estimate the RMSE with reasonable precision. The approach taken here is to include all products with at least 156 weeks (three years) of observationser tt16%, d6(here)8(

limited retail distribution represent a large percentage of category sales. Thus, consumers in the same market have very different c

References

Asker, John, "Measuring Advantages from Exclusive Dealing," working paper, April 2004a.

Asker, John, "Diagnosing Foreclosure Due to Exclusive Dealing," working paper, October 2004b.

Bergen, Mark, Shantanu Dutta, and Steven M. Shugan, "Branded Variants: A Retail Perspective," *Journal of Marketing Research* 33 (1996), pp. 9-19.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25 (1994), pp. 242-262.

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

Besanko, David, Jean-Pierre Dubé, and Sachin Gupta, "Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Retail Data," *Management Science* 49 (2003), pp. 1121-38.

Boatwright, Peter, Sanjay Dhar, and Peter E. Rossi, "The Role of Retail Competition, Demographics and Account Retail Strategy as Drivers of Promotional Sensitivity," *Quantitative Marketing and Economics* 2 (2004), pp. 169-90.

Brenkers, Randy, and Frank Verboven, "Liberalizing a Distribution System: The European Car Market," working paper, November 2004.

Bronnenberg, Bart J., Vijay Mahajan, and Wilfried R. Vanhonacker, "The Emergence of Market Structure in New Repeat-Purchase Categories: The Interpl(e Economih R)4bTncD [(Brenkr17 Tc rD)]TJ -0.001

Curhan, Ronald C., “The Relationship Between Shelf Space and Unit Sales in Supermarkets,” *Journal of Marketing Research* 9 (1972), pp. 406-12.

Draganska, Michaela and Dipak Jain, “Product-Line Length as a Competitive Tool,” *Journal of Economics and Management Strategy* 14 (2005), pp. 1-28.

Drèze, Xavier, Stephen J. Hoch, and Mary E. Purk, “Shelf Management and Space Elasticity,” *Journal of Retailing* 70 (1994), pp. 301-26.

Farris, Paul, James Olver, and Cornelis De Kluyver, “The Relationship Between Distribution and Market Share,” *Marketing Science* 2 (1989), pp. 107-28.

Gupta, Sachin, Pradeep Chintagunta, Anil Kaul, and Dick R. Wittink, “Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Level Data,” *Journal of Marketing Research* 33 (1996), pp. 383-98.

Hausman, Jerry A. and Gregory K. Leonard, “The Competitive Effects of a New Product Introduction: A Case Study,” *Journal of Industrial Economics* 50 (2002), pp. 237-63.

Hausman, Jerry, Gregory Leonard, and J. Douglas Zona, “Competitive Analysis with Differenciated Products,” *Annales D’Économie et de Statistique* 34 (1994), pp. 159-80.

Heide, Jan B., Shantanu Dutta, and Mark Bergen, “Exclusive Dealing and Business Efficiency: Evidence From Industry Practice,” *Journal of Law and Economics* 41 (1998), pp. 387-407.

Hoch, Stephen J., Byung-Do Kim, Alan L. Montgomery, and Peter E. Rossi, “Determinants of Store-Level Price Elasticity,” *Journal of Marketing Research* 32 (1995), pp. 17-29.

Kumar, V. and Robert P. Leone, “Measuring the Effect of Retail Store and Promotions on Brand and Store Substitution,” *Journal of Marketing Research* 25 (1988), pp. 178-85.

Lafontaine, Francine, “Agency Theory and Franchising: Some Empirical Results,” *RAND Journal of Economics* 23 (1992), pp. 263-83.

Lafontaine, Francine and Kathryn L. Shaw, “The Dynamics of Franchise Contracting: Evidence from Pane Data,” *Journal of Political Economy* 107 (1999), pp. 1041-80.

Lafontaine, Francine and Margaret Slade, “Retail Contracting: Theory and Practice,” *Journal of Industrial Economics* 45 (1997), pp. 1-25.

Nevo, Aviv, “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand,” *Journal of Economics and Management Strategy* 9 (2000), pp., 513-548.

Nevo, Aviv, “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica* 69 (2001), pp. 307-42.

Nevo, Aviv and Konstantinos Hatzitaskos, “Why Does the Average Price of Tuna Fall During Lent?,” NBER Working Paper No. 11572, August 2005.

Olver, James M. and Paul W. Farris, “Push and Pull: A One-Two Punch for Packaged Products,” *Sloan Management Review* 31 (1989), pp. 53-61.

Ornstein, Stanley I. and Dominique M. Hanssens, “Resale Price Maintenance: Output Increasing or Restricting? The Case of Distilled Spirits in the United States,” *Journal of Industrial Economics* 36 (1987), pp. 1-18.

Pesendorfer, Martin, “Retail Sales: A Study of Pricing Behavior in Supermarkets,” *Journal of Business* 75 (2002), pp. 33-66.

Petrin, Amil, “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy* 110 (2002), pp. 705-29.

Reibstein, David J. and Paul W. Farris, “Market Share and Distribution: A Generalization, A Speculation, and Some Implications,” *Marketing Science* 14 (1995), pp. 190-202.

Rhee, Hongjai and David R. Bell, “The Inter-Store Mobility of Supermarket Shoppers,” *Journal of Retailing* 78 (2002), pp. 225-37.

Shaffer, Greg, “Slotting Allowances and Resale Price Maintenance: A Comparison of Facilitating Practices,” *RAND Journal of Economics* 22 (1991), pp. 120-35.

Sullivan, Mary W., “Slotting Allowances and th

Table 1

Percent of Products, by Median Product Availability

Available in % of Stores:	Frozen Novelty (N=3,123)	Pasta (N=5,127)	Hot Dog (N=1,176)	Ice Cream (N=6,345)	Salad Dressing (N=6,436)	Average
Low Availability:						
0% to 10%	9%	26%	10%	8%	17%	14%
10% to 20%	8%	18%	6%	10%	14%	11%
20% to 30%	5%	7%	5%	7%	7%	6%
30% to 40%	5%	5%	3%	7%	5%	5%
40% to 50%	6%	4%	3%	8%	5%	5%
Sub-total	33%	60%	28%	40%	48%	42%
Intermediate Availability:						
50% to 60%	7%	4%	5%	9%	6%	6%
60% to 70%	9%	4%	6%	10%	7%	7%
70% to 80%	12%	5%	7%	10%	9%	9%
80% to 90%	15%	7%	13%	11%	11%	11%

40% to11% 11% 11%

Table 2

Monte Carlo Results

Table 4

Logit Estimates, Product Introduction Duration Model

A. Dependent Variable: Product is Available

	Frozen Novelty			Pasta			Hot Dog			Ice Cream			Salad Dressing		
	Est	SE		Est	SE		Est	SE		Est	SE		Est	SE	
Intercept	4.36	(0.39)	***	3.51	(0.23)	***	3.13	(0.54)	***	4.06	(0.30)	***	4.31	(0.22)	***
Time Elapsed	22.25	(3.92)	***	20.48	(2.52)	***	22.53	(6.20)	***	27.52	(3.17)	***	18.06	(2.20)	***
Time Elapsed^2	-103.23	(17.61)	***	-73.04	(11.06)	***	-88.44	(31.56)	***	-124.52	(14.07)	***	-83.15	(9.95)	***
Time Elapsed^3	159.96	(28.77)	***	99.69	(17.52)	***	128.00	(57.45)	**	190.60	(22.44)	***	133.64	(16.19)	***
Time Elapsed^4	-81.96	(15.53)	***	-48.42	(9.07)	***	-63.19	(33.54)	*	-96.52	(11.80)	***	-71.55	(8.63)	***
Product Line is Widely Available	-0.15	(0.25)		0.61	(0.17)	***	-1.64	(0.64)	**	0.19	(0.13)		0.72	(0.18)	***
Product Line Contains a Widely Available UPC	0.17	(0.28)		-0.05	(0.28)		2.04	(0.72)	***	-0.30	(0.19)		-0.40	(0.21)	*
Product Line Contains a Widely Available UPC of the Same Variety	-0.14	(0.33)		-0.81	(0.32)	**	-0.26	(0.73)		0.30	(0.62)		-0.20	(0.20)	
Product Line Contains a Widely Available UPC with the Same Package Size	0.21	(0.22)		-0.43	(0.27)		-0.47	(0.45)		0.30	(0.16)	*	-0.33	(0.15)	**
# of Observations	36,454			82,987			9,067			77,732			91,323		

B. Dependent Variable: Product is Widely Available

	Frozen Novelty			Pasta			Hot Dog			Ice Cream			Salad Dressing		
	Est	SE		Est	SE		Est	SE		Est	SE		Est	SE	
Intercept	-7.60	(0.50)	***	-7.20	(0.39)	***	-6.37	(0.74)	***	-5.19	(0.24)	***	-7.17	(0.27)	***
Time Elapsed	15.55	(3.58)	***	18.02	(5.15)	***	-5.09	(5.08)		20.14	(2.57)	***	18.53	(2.86)	***
Time Elapsed^2	-141.76	(26.69)	***	-129.76	(35.50)	***	-53.81	(37.75)		-183.36	(19.45)	***	-147.17	(20.62)	***
Time Elapsed^3	305.25	(63.45)	***	228.88	(78.58)	***	180.19	(89.22)	**	381.47	(45.75)	***	284.38	(47.17)	***
Time Elapsed^4	-194.25	(46.21)	***	-121.57	(53.77)	**	-140.62	(64.84)	**	-233.69	(33.23)	***	-165.34	(33.22)	***
Product Line is Widely Available	-0.33	(0.21)		0.60	(0.35)	*	0.39	(0.70)		-0.53	(0.12)	***	0.92	(0.15)	***
Product Line Contains a Widely Available UPC	0.23	(0.23)		1.76	(0.36)	***	0.21	(0.70)		0.63	(0.14)	***	0.39	(0.15)	**
Product Line Contains a Widely Available UPC of the Same Variety	-0.73	(0.31)	**	0.25	(0.27)		0.29	(0.42)		-0.01	(0.27)		0.48	(0.12)	***
Product Line Contains a Widely Available UPC with the Same Package Size	-0.09	(0.16)		0.76	(0.18)	***	0.27	(0.24)		-0.18	(0.10)	*	0.39	(0.10)	***
# of Observations	37,462			83,851			9,318			79,667			92,964		

Notes: The data sample is restricted to new products. Each observation corresponds to a particular UPC in a given retailer-city in a given week. Both models also include a set of dummy variables for the calendar month and a set of retailer-city fixed effects. Robust standard errors that are clustered by product are reported. Significance levels correspond to *=10%, **=5%, ***=1%. For clearer presentation, "time elapsed" is rescaled by dividing by 132 weeks, so that this variable takes values between 0 and 1.

Table 5

Spell Length Predicted Probabilities for New Product Introductions

A. Outcome in First Year Following Product Introduction

	Frozen Novelty (N=1,340)	Pasta (N=1,631)	Hot Dog (N=358)	Ice Cream (N=2,581)	Salad Dressing (N=2,634)	Average
			% of Products			
% Widely Available	17%	27%	49%	51%	26%	34%
% Available, But Not Widely Available	34%	57%	27%	34%	41%	39%
% Discontinued	49%	17%	24%	15%	33%	28%
Total	100%	100%	100%	100%	100%	100%
			Weeks in First Year with Limited Distribution			
Average weeks	25.3	36.1	21.4	25.5	30.6	27.8
Median weeks	15.0	52.0	10.0	16.0	33.0	25.2

B. Duration, Conditional on Becoming Widely Available in First Year

	Frozen Novelty	Pasta	Hot Dog	Ice Cream	Salad Dressing	Average
			% of Products			
1-4 weeks	29%	18%	43%	26%	19%	27%
5-8 weeks	26%	20%	23%	26%	21%	23%
9-12 weeks	19%	16%	13%	18%	19%	17%
13-26 weeks	20%	33%	15%	24%	32%	25%
27-39 weeks	4%	10%	4%	4%	7%	6%
40-52 weeks	2%	3%	2%	1%	2%	2%
Total	100%	100%	100%	100%	100%	100%
Average weeks	10.1	14.2	8.6	10.3	12.8	11.2
Median weeks	8.0	11.0	5.0	8.0	10.0	8.4

C. Duration, Conditional on Being Discontinued in First Year

	Frozen Novelty	Pasta	Hot Dog	Ice Cream	Salad Dressing	Average
			% of Products			
1-4 weeks	20%	21%	28%	25%	17%	22%
5-8 weeks	16%	18%	19%	16%	15%	17%
9-12 weeks	11%	12%	11%	9%	10%	11%
13-26 weeks	19%	23%	18%	14%	22%	19%
27-39 weeks	13%	13%	11%	11%	16%	13%
40-52 weeks	21%	13%	13%	24%	20%	18%
Total	100%	100%	100%	100%	100%	100%
Average weeks	20.1	17.5	16.1	20.1	21.1	19.0
Median weeks	14.0	12.0	9.0	12.0	16.0	12.6

Notes: N = number of products, which is defined as a UPC in a given retailer-city.

Table 6

Table 7

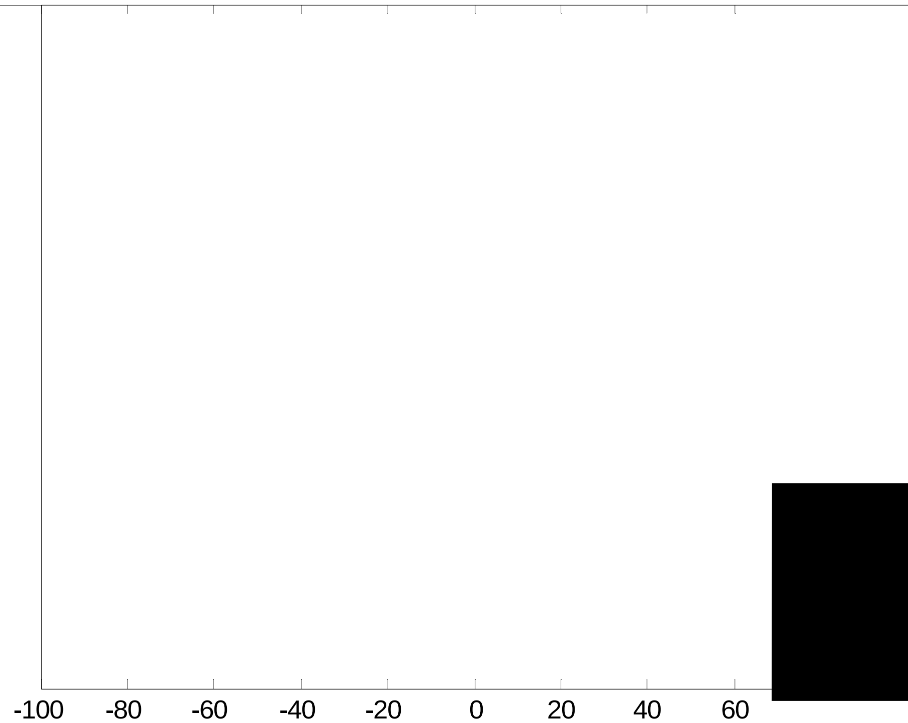
RMSE of Log Product Distribution

RMSE:	Frozen Novelty (N=178)	Pasta (N=349)	Hot Dog (N=91)	Ice Cream (N=336)	Salad Dressing (N=335)	Average
	% of Products					
Low Variability:						
0% to 5%	3%	21%	24%	6%	9%	13%
5% to 10%	22%	36%	25%	15%	18%	23%
Sub-total	25%	56%	49%	22%	26%	36%
Intermediate Variability:						
10% to 15%	25%	19%	20%	30%	21%	23%
15% to 20%	16%	12%	14%	19%	19%	16%
Sub-total	40%	31%	34%	49%	39%	39%
High Variability:						
20% to 25%	15%	8%	7%	13%	15%	11%
>25%	20%	5%	10%	17%	20%	14%
Sub-total	34%	13%	16%	30%	34%	26%
Mean RMSE	18%	11%	12%	17%	17%	15%
Median RMSE	15%	9%	10%	15%	16%	13%

Notes: N = number of products, which is defined as a UPC. The table reports the root mean square error from regressions using log ACV as the dependent variable. A set of retailer-city and time fixed effects are employed, along with a retailer-city specific quadratic time trend. Each column sums to 100%.

Figure 1

Histogram of the Cross-Price Elasticity Percent Bias for Products in the Hot Dog Category



[redacted]

[redacted]