I. Introduction

The usefulness of scanner data for analyzing the retail sector is widely seen as a "success story" by both academics and industry participants (Bucklin and Gupta 1999). It remains an open question, however, whether aggregate-level data can reliably be used to estimate the demand for a set of products, or if store-level data is required. Although previous research shows that demand estimates based on aggregate data are biased when stores are heterogeneous, only partial solutions to the aggregation bias problem have been developed thus far.

We propose a methodology for avoiding aggregation bias that allows inter-store heterogeneity to be explicitly controlled for with aggregate data. This is accomplished by exploiting information regarding the distribution of store characteristics, information that is only partially utilized in extant aggregate demand models. Our approach is highly practical since it relies solely on standard scanner data of the type produced by the major vendors, ACNielsen and Information Resources, Inc. (IRI).

Throughout this paper, "aggregate-level" refers to data where the sales from multiple stores are combined. Examples of aggregate datasets include city-level data (e.g., all supermarkets in Chicago), and city-chain data (e.g., all of Jewel's supermarkets in Chicago).[1] Researchers who lack access to store-level data must depend on these types of datasets to estimate the demand for a set of products. Unfortunately, aggregate-level scanner datasets do not report the marketing-mix characteristics of products at each store, such as their price and promotional activity. Unable to model inter-store heterogeneity with such data, researchers have employed a representative store paradigm where each consumer faces the average price and promotion level across all stores. Recent examples include Nevo (2000a), Hausman and Leonard (2002), Cotterill and Samson (2002), and Perloff and Ward (2003). Modeling aggregate

[1] A confidentiality agreement with ACNielsen prohibits retailer names from being revealed. This example does not indicate whether the dataset employed contains the Jewel supermarket chain in Chicago.

demand in this simplified manner comes at a high cost, however. The aggregation bias literature demonstrates that when stores have heterogeneous marketing-mix strategies, the representative store model produces biased demand univariate distribution of each product's promotional activity, information that is reported in the scanner datasets produced by the two major vendors.

We demonstrate the advantages of our proposed methodology by estimating the demand for super-premium ice cream. Using a random coefficients logit demand model, we not only show that our framework produces sensible results, but estimates that are measured more precisely than in the standard, representative store model. In contrast, the traditional model produces implausible estimates due to aggregation bias. These findings are corroborated by Monte Carlo analysis that shows our promotional disaggregation approach substantially outperforms the representative store framework.

The paper is organized as follows. Section two reviews the methods that have previously been developed to avoid aggregation bias. Section three presents a consumer demand model that accounts for marketing-mix heterogeneity across stores, and then uses this framework to estimate the demand for super-premium ice cream. In section four, we employ Monte Carlo analysis to compare our disaggregated promotion framework to the standard representative store model. Section five concludes.

II. Aggregation Bias

Estimating demand with aggregate data often leads to model mis-specification, or "aggregation bias." This section reviews two previously developed solutions for avoiding aggregation bias. Since this problem is widely known, we do not detail why aggregation bias occurs. For a detailed consideration of this issue, Theil (1954) and Krishnamurthi et al. (1990) analyze the linear model; Lewbel (1992), Christen et al. (1997), and Chung and Kaiser (2000) analyze the constant elasticity model; and Allenby and Rossi (1991) and Krishnamurthi et al. (2000) analyze the logit model.

Link (1995) argues that data aggregation across stores with heterogeneous marketing activity is the most significant source of bias in practical applications. Link suggests that aggregation bias be avoided by employing data that has been aggregated across stores with homogeneous marketing activity.[2] However, even if one obtains data that is aggregated across stores where a product's own promotions are homogenous, heterogeneity in the promotions of competing products may remain. Thus, Link's approach does not account for aggregation bias in cross-product effects. A further limitation is that it requires demand for each product to be separately estimated, since it is possible to disaggregate by promotional activity for only one product at a time. This prevents it from being applied to certain frameworks, such as the random coefficients logit model where the demand for each product is jointly estimated (Nevo 2000b).

Christen et al. (1997) propose a methodology to "de-bias" demand estimates based on aggregate data. First, demand is estimated using simulated store-level data that has been aggregated across stores. The average difference between the true and estimated parameters from the simulation is then added to the estimates from an empirical application, to de-bias the results. It can be difficult to estimate the magnitude of aggregation bias reliably as one may have insufficient information to calibrate the simula

Data

We utilize supermarket scanner data provided by ACNielsen for the super-premium ice cream category. The dataset separately reports weekly sales for 14 city-chain combinations for the period December 1998 to June 2001 (132 weeks). However, to avoid complications involving entry into certain geographic areas, only a subset of the data is used; we analyze the last 80 weeks of data for the 11 city-chain combinations where the same four brands comprise the entire category. To comply with a confidentiality agreement with ACNielsen, they are referred to as Brand A, B, C, and D.

The data separately reports unit and dollar sales for four mutually exclusive levels of promotional activity $m \in M$, where M = {"No Promotion," "Display Only," "Feature Only," "Feature & Display"}. A "Feature" is a print advertisement, such as in a promotional circular, while a "Display" is a secondary sales location within a store used to draw special attention to a given product. The demand specification presented below details the conditions under which consumer demand can be added up across the subset of stores where a product has a given type of promotional activity, so that it can be consistently estimated with aggregate data. The required conditions are less restrictive than those needed to perform an equivalent aggregation exercise using a representative store model.

In the super-premium ice cream category each brand's UPCs represent a different flavor, with a particular flavor rarely available for more than one brand (e.g., "Chunky Monkey" is available only for the Ben & Jerry's brand).[3] The large number of idiosyncratic flavors limits the usefulness of this characteristic for estimating substitution patterns. Other meaningful characteristics are common across UPCs for a given brand; each brand's UPCs share the same brand image and, within any given store, they are identically priced and promoted. Below, we

[3] In addition, UPCs vary by package size. Most brands of super-premium ice cream are available only in pint-sized containers, however, with larger package sizes representing a small fraction of category sales. We therefore omit them from the analysis by restricting the dataset to pint-sized cartons.

develop a product-level demand specification. Nonetheless, since the control variables employed vary only by brand, this specification simplifies to a brand-level demand model.

<u>Disaggregated Promotion Model</u>

The following details the random coefficients logit demand model employed in the empirical analysis. In every time period *t*, each consumer *i* purchases that item which generates the highest utility. The choices are the set of currently available products J_t or the "outside good." We normalize the utility derived from purchasing the outside good to a mean utility of zero, $U_{i0t} \quad {}_{i0t}$, where ${}_{i0t}$ is i.i.d. Type I Extreme Value. For the remaining choices, consumer *i*'s utility for product *j* during week *t* is determined by its promotional activity $m_{ijt} \quad M$, price p_{ijt}, a set of product characteristics X_{ijt} that has an associated vector of random coefficients ${}_{i}$, a set of additional controls Z_{jt}, and an i.i.d. error term ${}_{ijt}$ that is distributed Type I Extreme Value.[4]

$$(3.1) \quad U_{ijt} \qquad {}^{m_{ijt}} \qquad {}^{m_{ijt}} p_{ijt} \quad X_{ijt} \; {}_{i} \quad Z_{jt} \qquad {}_{ijt}$$

Product characteristics X_{ijt} include a set of dummy variables for each brand, price p_{ijt}, and dummy variables for "Display Only," "Feature Only" and "Feature & Display." Control variables Z_{jt} consist of brand fixed effects for each city-chain combination, a fourth order time trend, the number of products availa

The model accommodates heterogeneity in consumer preferences through random coefficients $_i$. We assume $_i$ is mean-zero and i.i.d. Normally distributed with a block diagonal variance matrix $V \quad \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}$. Denote the probability distribution function of $_i$ by $(\ _i;V)$. V_1 corresponds to the brand dummy variables contained in X_{ijt} (the fixed characteristics), while V_2 corresponds to the remaining price and promotion variables (the variable characteristics). We place no restrictions on V_1 and V_2

(3.3) $$U_{ijt} \quad {}^{m_j(g_{it})} \quad {}^{m_j(g_{it})} p_{jt}^{m_j(g_{it})} \quad X_{jt}(g_{it}) \;{}_{i} \quad Z_{jt} \quad {}_{ijt}$$

Apart from heterogeneity in price and promotional activity, all stores are identical. In addition, we assume each consumer is randomly matched to a store. This allows us to integrate over the distribution of random coefficients $_i$ for the subset of consumers who visit a given store type *g*, even though aggregate scanner datasets contain no information regarding individual stores or consumers.

Each consumer *i*oduct/TT1 1 Tf 0 Tc 0 Tw 12.015 0 0 12.015 122.028 -2 72 Tm (m)TjTT0 1 Tf 0.0005 T

$$\text{(3.5)} \quad \hat{q}^{g}_{bt} \qquad \underset{j \quad J_t : b_j \quad b}{} \hat{q}^{g}_{jt} \quad Q_{\quad gt} \; \frac{\underset{j \quad J_t : b_j \quad b}{} e^{[\quad m_j(g) \qquad m_j(g)\, p_{jt}^{m_j(g)} \quad X_{jt(g)} \;\, i \quad Z_{jt} \;]}}{1 \qquad \underset{\tilde{b} \quad B \quad j \quad J_t : b_j \quad \tilde{b}}{} e^{[\quad m_j(g) \qquad m_j(g)\, p_{jt}^{m_j(g)} \quad X_{jt(g)} \;\, i \quad Z_{jt} \;]}} \; (\;_i ; V) d \;\; _i$$

All variables in (3.5) that have a product *j* subscript are identical across products with the same brand *b*. To simplify notation we therefore replace each *j* subscript with a b_j subscript. Equation (3.5) reduces to the following, where N_{bt} denotes the number of products available in time *t* that are part of brand *b*'s product line.

$$\text{(3.6)} \quad \hat{q}^{g}_{bt} \quad Q_{\quad gt} \; \frac{N_{bt} e^{[\quad m_b(g) \qquad m_b(g)\, p_{bt}^{m_b(g)} \quad X_{bt(g)} \;\, i \quad Z_{bt} \;]}}{1 \qquad \underset{\tilde{b} \quad B}{} N_{\tilde{b}t} e^{[\quad m_{\tilde{b}}(g) \qquad m_{\tilde{b}}(g)\, p_{\tilde{b}t}^{m_{\tilde{b}}(g)} \quad X_{\tilde{b}t(g)} \;\, i \quad Z_{\tilde{b}t} \;]}} \; (\;_i ; V) d \;\; _i$$

To account for our transformation from a product- to a brand-level demand model, we update the definition of store type *g*. We now define *g* as the set of promotional activity across the four brands. Since there are four brands and four types of promotional activity, *G* contains 4^4 = 256 unique store types. This simplification is possible since, as we discussed earlier, in the super-premium ice cream category each brand's entire product line is identically promoted within any given store.

Calculation of $\hat{q}^{g}_{bt}$ in equation (3.6) requires integration over the random coefficients $_i$. One way to do so is to generate a random sequence of draws $\{ \; ^l \}_{l \;\; 1}^{L}$ that are mean-zero and i.i.d. Normally distributed with variance matrix *V*. We then approximate equation (3.6) as follows.

(3.7)

Generating $\{\ ^{l}\}_{l\ \ 1}^{L}$ using a Halton sequence is a more efficient means of calculating $\hat{q}_{bt}^{g}$ (Train 1999). Since the Halton sequence produces values that are more smoothly distributed over the support of the Normal distribution than would occur under random sampling, we can choose a much smaller value for L and still obtain accurate results. Nonetheless, we set 000

(3.9) $\tilde{U}_{ijt} \quad E(U_{ijt} \mid p_{ijt} \quad p_{jt}, \; {}_{i}, X_{jt}, Z_{jt}) \quad {}_{ijt} \quad \bar{}_{jt} \quad \bar{}_{jt}\, p_{jt} \quad X_{jt} \; {}_{i} \quad Z_{jt} \quad {}_{ijt},$

where $\bar{}_{jt} \quad {}_{m \; M} \quad {}^{m} \; {}^{m}_{jt}$ and $\bar{}_{jt} \quad {}_{m \; M} \quad {}^{m} \; {}^{m}_{jt}$

Utility function (3.9) implies the following characterization of predicted unit sales for product *j* in week *t*.

(3.10)

stringently assumes that every consumer observes the same price. The disaggregated promotion model recognizes that if a product is on promotion in 20% of stores, then 20% of consumers observe the promotion and 80% do not. In contrast, the representative store model assumes that consumers observe the average promotional activity across all stores (i.e., everyone observes a "partial" promotion). To summarize, the representative store framework ignores inter-store heterogeneity, averaging over differences in price and promotional activity. This approach is problematic since previous research demonstrates it leads to aggregation bias. In contrast, our disaggregated promotion framework explicitly models heterogeneous store types.

<u>Data Requirements</u>

Estimation of the disaggregated promotion model requires only standard scanner data of the type produced by the major vendors, ACNielsen and IRI. Typically, such data separately reports dollar and unit sales for four (mutually exclusive) types of promotional activity: "No Promotion," "Display Only," "Feature Only," and "Feature & Display." Given assumption (3.2), price is calculated as dollar sales for promotion *m* divided by unit sales for that promotion.

Scanner data reports information regarding product and promotional distribution through a variable known as "All Commodity Volume," or ACV. ACV_{jt} is the percentage of total sales, across all product categories, accounted for by those stores that carry product *j* in week *t*. This represents the percentage of stores that distribute a particular item. Similarly, ACV_{jt}^{m} is the fraction of stores where product *j* has promotional activity *m*. Note that the percentage of stores that carry product *j* is the sum of its promotional distribution: $ACV_{jt} \quad \underset{m \;\; M}{} ACV_{jt}^{m}$.

We use these distribution measures to calculate two variables. First, the model requires N_{bt}, the number of brand *b*'s products that are available in time *t*. We also need $^{m}_{bt}$, the fraction of stores where brand *b* has promotional activity *m*. Standard scanner data contains sufficient information to construct both N_{bt} and $^{m}_{bt}$. To calculate the number of products contained within brand *b*'s product line, we add up each product's ACV: $N_{bt} \quad \underset{j \;\; J_t : b_j \;\; b}{} ACV_{jt}$.

We then add up the fraction of stores where each of brand *b*'s products has promotional activity *m*: $N_{bt}^{m} \quad \underset{j \quad J_t : b_j \quad b}{} ACV_{jt}^{m}$. The percentage of stores where brand *b* has promotion *m* is calculated as ${}_{bt}^{m} \quad \frac{N_{bt}^{m}}{N_{bt}}$.

While each brand's univariate promotional distribution $\{ {}_{bt}^{m} \}_{m \quad M}$ is calculated in this manner, the joint distribution of each brand's promotions $\{ {}_{t}^{g} \}_{g \quad G}$ is not reported in aggregate scanner datasets. Additional model restrictions must be imposed in order to estimate the joint promotional distribution, since a continuum of joint distributions is generally possible given a set of univariate distributions.

There is a special case, however, where the joint promotional distribution is uniquely determined by each brand's univariate distribution. Define brand *b* as having heterogeneous promotions in week *t* when its promotional activity varies across the stores aggregated in the data. That is, when $\underset{m \quad M}{} 1 \quad {}_{bt}^{m} \quad {}_{0} \quad 1$. The joint promotional distribution is uniquely determined by each brand's univariate distribution if no more promotional distribution is uniquely determined by the observed marginal distributions. Alternatively, to apply the model to the 15% of observations where the joint distribution is not known, we have to make an additional assumption about functional form to get identification. Specifically, we assume the joint promotional distribution can be constructed from a copula of the marginal distributions. The data is used to estimate the single parameter of the copula jointly with the other model parameters.

Deciding between these two approaches involves the familiar bias-variance tradeoff. The obtained estimates will be more precise if one imposes additional structure that allows the model to be estimated from the full data sample. However, they may be biased if the employed assumptions are invalid. We believe that, on balance, the benefit of exploiting the entire data sample outweighs the cost of imposing the additional model structure detailed below. Of course, those who believe the benefit does not outweigh the cost can instead estimate the model using the subset of the data where the joint promotional distribution is known.

We rely on the following framework. Let each retail chain be composed of a continuum of stores. Brand *b*'s promotional activity in store *s* during week *t* is determined by latent variable ${}_{bst}$, which has a standard Normal distribution. This variable is used to assign brand *b*'s promotional activity in store *s*. We assume brand *b*'s promotional activity is weakly increasing in ${}_{bst}$ based on the following rank order of promotional activity, from lowest to highest: "No Promotion," "Display Only," "FeatuNon Tc -0. in sNTd [ng

We then use a Gaussian copula to specify the joint distribution of each brand's latent variable $_{bst}$. Define $_{st}$ $\{ _{bst}\}_{b \; B}$, where vector $_{st}$ is mean-zero and i.i.d. Normally distributed with variance matrix . Denote the probability distribution function of $_{st}$ by $(_{st};)$. To minimize the number of estimation parameters, we assume has identical off-diagonal elements $[0,1]$ and unit values along the main diagonal.[8]

Parameter represents retailer strategy regarding how products are jointly promoted across stores. Retailers independently set each brand's promotional activity when equals zero. As increases, the promotional activity of competing brands becomes more positively correlated. That is, in stores where a retailer chooses a high level of promotional activity for one brand, for larger values of it more frequently chooses a high level of promotion for the other brands in those stores.

This framework provides sufficient structure to calculate the joint promotional distribution $\{ _t^g \}_{g \; G}$. To calculate $_t^g$ for each $g \; G$, we numerically integrate $(_{st};)$ over the range of values where the promotional activity of each brand $b \; B$ equals $m_b(g)$.

(3.12) $$_t^g \quad \int_{_t^g} (_{st};) d \, _{st} \, ,$$

where $$_t^g \quad \{ _{st} : \; _{bst} \quad [\quad ^{1}(\sum_{m \; M : m \; m_b(g)} {}_{bt}^{m}), \quad ^{1}(\sum_{m \; M : m \; m_b(g)} {}_{bt}^{m})], \quad b \; B \}$$

To summarize, this framework uses each brand's univariate promotional distribution to choose a joint promotional distribution from a family characterized by estimation parameter . We estimate jointly with the other demand parameters via maximum likelihood (see the following subsection). As discussed below, is identified by how variation in this parameter

[8] Parameter does not vary over time and is identical across retailers. We make this simplifying assumption since only 15% of the dataset's observations identify the joint promotional distribution. A more flexible specification can be employed in situations where it is practical to do so.

impacts predicted market shares. It is not possible to estimate prior to solving for the other demand parameters, since predicted market shares cannot be computed without them.

However, one can use the following two-stage estimation procedure to solve for after using a subset of the data to estimate the equals observed quantity sold. This specification is theoretically appealing since the error structure is integrated within the utility-based demand model.[9]

This approach has two drawbacks, however. First, the proposed inversion method is computationally intensive. Second, it requires a strong belief that the “correct” model is being employed; Berry discusses how his inversion method is sensitive to model mis-specificatiodr341Cn variance as a second order polynomial in ${}^{m}_{bt}$, the fraction of stores where brand *b* has promotional activity *m*.

$$(3.13)\quad Var({}^{m}_{bt}) \quad e^{\;0 \quad 1\,{}^{m}_{bt} \quad 2({}^{m}_{bt})^{2}}$$

Denote the probability distribution function of ${}^{m}_{bt}$ by $\;({}^{m}_{bt}, Var({}^{m}_{bt}))$

the percentage of stores, for each level of promotional activity other than "No Promotion." This is due to two distinct effects. First, promotions lead to an outward shift in the demand curve for a given brand. Promotional activity is also associated with a price reduction, with approximately 10% lower prices when on "Display Only," and 30% lower when on "Feature Only" or "Feature & Display." These promotional price reductions are a second factor leading to increased sales.

Table 2 presents parameter estimates for the disaggregated promotion model. Price coefficient ^{m} and intercept ^{m} increase (in absolute value) in the level of promotion *m*, with "No Promotion" the lowest promotional activity, "Display Only" and "Feature Only" intermediate promotions, and "Feature & Display" the highest type of promotional activity. The net impact of these two parameter changes is that a promotion unaccompanied by a price reduction leads to only a small, positive increase in consumer utility (and therefore sales). However, since promotions make consumer utility a steeper function of price, a price reduction accompanied by promotional activity has greater impact than the same price reduction and promotion when separately undertaken.

Table 2 also presents parameter estimates for the representative store model. The parameters for "Display Only" and "Feature & Display" are imprecisely estimated. Table 3 reveals why this is the case. Retailers typically employ these types of promotions in only a small fraction of stores in any given week. For example, when Brand A is on "Display Only" in at least one store in a city-chain, on average 2.5% of stores promote Brand A in this manner. Similarly, on average only 8.9% of stores have a "Feature & Display." The representative store model is unable to isolate the impact of "Display Only" or "Feature & Display" using data aggregated with promotions that are more prevalent. In contrast, the disaggregated promotion model estimates these effects quite precisely.

In addition to the demand estimates presented in Table 2, the disaggregated promotion model has an additional parameter . Recall that this parameter is used to estimate the joint distribution of each brand's promotional activity. We obtained the corner solution 1. As a robustness check we re-estimated the model assuming 0.[11] We obtained similar results for the other demand parameters. This insensitivity to the value of does not imply, however, that aggregation bias is not a problem. As discussed earlier, the value of affects only those 15% of weeks where at least two brands have heterogeneous promotions. In contrast, aggregation bias affects the results of the representative store model even when only one brand has heterogeneous promotions. In this case, only one joint promotional distribution can arise given each brand's univariate distribution. The representative store model ignores this information, and instead assumes every consumer observes the average promotional activity across stores. It is much more common for retail chains to promote a single brand than two or more brands at the same time. As such, the disaggregated promotion model requires that we estimate the joint distribution of promotions for a subset of those weeks where at least one brand has heterogeneous promotions across stores. This is why the demand estimates produced by our model are similar regardless of whether 0 or 1, even though aggregation bias significantly impacts the results of the representative store framework.

The first set of estimates in Table 4 presents each brand's own-price elasticity for each type of promotional activity. This is followed by the matrix of cross-price elasticity estimates, calculated when each brand is not on promotion. The third set of results reports the impact of each brand's own promotional activity relative to "No Promotion." All three sets of estimates are evaluated at each brand's average price for the given level of promotion, and are calculated assuming the other brands are not on promotion. This implies the promotional effects shown in the third set of results report the combined effect of being on promotion and undergoing the average price reduction for that promotion.

[11] We also considered the following alternative framework. We let $\{ {}^{g}_{t} \}_{g \; G}$ be a weighted average of two distributions: the distribution that arises when 0 and the distribution when 1. Using this specification, we obtain the same joint distribution as before, where each brand's promotions are positively correlated to the maximum possible extent.

As mentioned earlier, one key difference between the two sets of results is that the estimates for “Display Only” and “Feature & Display” are imprecisely measured in the representative store model. A second, more critical shortcoming is that the promotional effects in the representative store model are implausibly large. For example, while “Display Only” increases Brand A’s sales by 53.1% in the disaggregated promotion model, the representative store model predicts an enormous 1914.5% increase. The magnitude of this effect is not a result of imprecise estimates, since the standard error is “only” 315.7%. The representative store model produces similarly implausible estimates for other brands and types of promotions. This finding is consistent with previous research that concludes data aggregation across stores with heterogeneous promotional activity often leads to overestimation of own-brand promotional effects (Link 1995, Christen et al. 1997).

IV. Monte Carlo Analysis

The previous section demonstrates that the disaggregated promotion model generates reasonable demand estimates, while the representative store model does not. Nonetheless, it is impossible to state that the former model is superior without knowing the true parameter values. Therefore, this section uses Monte Carlo analysis to study differences between the two models, specifically whether the poor performance of the representative store model results from inadequate control of promotional heterogeneity across stores. We simulate data using the control variables from the super-premium ice cream data in conjunction with the parameter estimates for the disaggregated promotion model. The constructed data is then used to estimate the disaggregated promotion and representative store models. Since the representative store model generates imprecise results, we must employ a large number of Monte Carlo simulations to calculate accurately the average difference between the true and estimated values. The high computational burden of estimating the random coefficients logit model makes doing so impractical.

We therefore conduct this anal

The disaggregated promotion model requires we estimate an additional parameter ,

which determines the joint distribution of promotions as a function of each brand's univariate distribution. Table 6 reports the histogram of the estimates from the Monte Carlo simulations discussed above, where the true value of equals one. The table al before, a failure to model intra-promotional price heterogeneity leads only to minor bias; across the various estimates, the average percentage difference between the true and estimated values is never greater than 6.8%, and is generally much smaller. The representative store model continues to perform worse. As before, own-price elasticities and promotional effects are imprecisely estimated for “Display Only” and “Feature & Display.” In addition, the average impact of these promotions is quite different from the true value. This comparison demonstrates that the disaggregated promotion model is a dramatic improvement over the representative store framework, and can be successfully applied even when there is significant price variation across stores with the same promotional activity.

V. Conclusion

Demand estimation using aggregate data often leads to biased results. However, only limited solutions for avoiding aggregation bias currently exist. They either have informational requirements that go beyond what is typically available, or fail to fully control for promotional heterogeneity across stores. Due to these shortcomings, practitioners continue to rely on representative store aggregate demand models that ignore inter-store promotional heterogeneity, and which are inconsistent with adding up from consumer-level demand. Previous research demonstrates these are the primary factors leading to aggregation bias.

We show how to avoid these leading determinants of aggregation bias. Our framework generalizes beyond the representative store paradigm by explicitly modeling heterogeneous store types. An aggregate demand model consistent with store-level heterogeneity is constructed by adding up demand across each type of store. This formulation requires the fraction of stores of each type, which we show how to estimate using information included in the scanner datasets produced by the major vendors, ACNielsen and IRI.

The presented empirical application demonstrates how to apply our proposed methodology to extant aggregate demand models. We not only show how to avoid aggregation bias, but also obtain results that are more precisely estimated. This is confirmed by Monte Carlo

Krishnamurthi, Lakshman, Raja Selvam, and Michaela Draganska (2000), "Inference Bias in Cross-Sectional Aggregation," Working Paper.

Lewbel, Arthur (1992), "Aggregation with Log-Linear models," *Review of Economic Studies*, 633-642.

Link, Ross (1995), "Are Aggregate Scanner Data Models Biased?," *Journal of Advertising Research*, RC 8-12.

Nevo, Aviv (2000a), "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal IndustTndy-to-Eat

Table 1

Summary Statistics

Brand A	
No Promotion	Display Only

Table 2

Parameter Estimates

Model 1: Disaggregated Random Coefficients Logit Model

Mean Coefficients:

	Intercept	Price
No Promotion		**-0.61**
		(0.05)
Display Only	**0.60**	**-0.79**
	(0.42)	(0.12)
Feature Only	**0.83**	**-1.13**
	(0.28)	(0.09)
Feature & Display	**2.40**	**-1.34**
	(0.37)	(0.13)

Standard Deviation of Random Coefficients:

Price	**0.13**
	(0.08)
Display Only	**1.01**
	(0.53)
Feature Only	**1.71**
	(0.18)
Feature & Display	**1.22**
	(0.42)

Model 2: Standard Random Coefficients Logit Model

Mean Coefficients:

	Intercept	Price
No Promotion		**-0.59**
		(0.04)
Display Only	**-0.60**	**1.22**
	(2.77)	(0.91)
Feature Only	**1.07**	**-1.03**
	(0.27)	(0.09)
Feature & Display	**6.90**	**-2.44**
	(1.58)	(0.61)

Standard Deviation of Random Coefficients:

Price	**0.06**
	(0.10)
Display Only	**1.29**
	(0.64)
Feature Only	**0.96**
	(0.28)
Feature & Display	**0.04**
	(0.87)

Notes: Standard errors are reported in parentheses. Model 1: N=4,332, log-likelihood=-335.87, and RMSE=.34. Model 2:

Table 3

Average Percentage of Stores on Promotion

	No Promotion	Display Only	Feature Only	Feature & Display
Brand A	93.0%	2.5%	69.9%	8.9%
Brand B	98.2%	4.9%	87.0%	12.6%
Brand C	90.2%	2.3%	78.5%	11.1%
Brand D	99.4%	5.6%	89.9%	12.0%

Notes: For each calculation, the data sample is restricted to those observations where at least one store has the given type of promotional activity for that brand.

Table 4

Estimated Own- and Cross-Brand Effects

Disaggregated Random Coefficients Logit Model

Own-Price Elasticity by Promotion

	No Promotion	Display Only	Feature Only	Feature & Display
Brand A	**-1.62**	**-1.87**	**-1.84**	**-2.27**
	(0.07)	(0.24)	(0.15)	(0.23)
Brand B	**-1.66**	**-1.97**	**-1.90**	**-2.29**
	(0.06)	(0.24)	(0.15)	(0.22)
Brand C	**-1.56**	**-1.81**	**-1.72**	**-2.22**
	(0.07)	(0.23)	(0.14)	(0.22)
Brand D	**-1.81**	**-2.32**	**-2.15**	**-2.69**
	(0.08)	(0.28)	(0.18)	(0.25)

Cross-Price Elasticities

In response to a price increase by:

	Brand A	Brand B	Brand C	Brand D
Brand A	**-1.62**	**0.07**	**0.13**	**0.02**
	(0.07)	(0.01)	(0.03)	(0.00)
Brand B	**0.21**	**-1.66**	**0.16**	**0.03**
	(0.03)	(0.06)	(0.03)	(0.01)
Brand C	**0.13**	**0.05**	**-1.56**	**0.02**
	(0.03)	(0.01)	(0.07)	(0.00)
Brand D	**0.16**	**0.06**	**0.14**	**-1.81**
	(0.03)	(0.01)	(0.03)	(0.08)

Own-Brand Promotional Effects

	Display Only	Feature Only	Feature & Display

Standard Random Coefficients Logit Model

Own-Price Elasticity by Promotion

	No Promotion	Display Only	Feature Only	Feature & Display
Brand A	**-1.75**	**0.76**	**-1.88**	**-3.24**
	(0.07)	(0.62)	(0.16)	(0.77)
Brand B	**-1.66**	**1.30**	**-1.76**	**-3.14**
	(0.07)	(1.03)	(0.15)	(0.68)
Brand C	**-1.60**	**0.89**	**-1.72**	**-3.20**
	(0.07)	(0.69)	(0.16)	(0.78)
Brand D	**-1.85**	**1.54**	**-2.05**	**-4.30**
	(0.09)	(1.18)	(0.17)	(0.94)

Cross-Price Elasticities

In response to a price increase by:

	Brand A	Brand B	Brand C	Brand D
Brand A	**-1.75**	**0.04**	**0.07**	**0.01**
	(0.07)	(0.01)	(0.02)	(0.00)
Brand B	**0.11**	**-1.66**	**0.10**	**0.02**
	(0.03)	(0.07)	(0.03)	(0.01)
Brand C	**0.07**	**0.03**	**-1.60**	**0.01**
	(0.02)	(0.01)	(0.07)	(0.00)
Brand D	**0.09**	**0.04**	**0.08**	**-1.85**
	(0.02)	(0.02)	(0.02)	(0.09)

Own-Brand Promotional Effects

	Display Only	Feature Only	Feature &

Table 5

Monte Carlo Results

Table 6

Notes: For each value of , the table reports histograms from 5,000 Monte Carlo simulations.

Table 7

Monte Carlo Results, Allowing for Intra-Promotional Price Heterogeneity

	Average Percent Difference Disaggregated Logit Model				Average Percent Difference Standard Logit Model			
	Own-Price Elasticity by Promotion				**Own-Price Elasticity by Promotion**			
	No Promotion	Display Only	Feature Only	Feature & Display	No Promotion	Display Only	Feature Only	Feature & Display
Brand A	**2.88%**	**4.27%**	**5.09%**	**5.74%**	**2.63%**	**7.77%**	**4.22%**	**-2.98%**
	(3.32%)	(4.63%)	(5.29%)	(5.80%)	(3.97%)	(84.30%)	(7.58%)	(42.34%)
Brand B	**2.79%**	**4.10%**	**4.71%**	**5.07%**	**2.55%**	**8.30%**	**3.94%**	**-2.70%**
	(3.32%)	(4.61%)	(5.22%)	(5.64%)	(3.97%)	(84.60%)	(7.49%)	(41.46%)
Brand C	**2.88%**	**4.27%**	**5.09%**	**5.72%**	**2.63%**	**7.94%**	**4.24%**	**-2.84%**
	(3.32%)	(4.64%)	(5.31%)	(5.83%)	(3.97%)	(84.52%)	(7.60%)	(42.60%)
Brand D	**2.77%**	**4.06%**	**4.61%**	**4.86%**	**2.53%**	**8.74%**	**3.86%**	**-2.39%**
	(3.32%)	(4.61%)	(5.23%)	(5.64%)	(3.97%)	(85.04%)	(7.50%)	(41.62%)
	Cross-Price Elasticities In response to a price increase by:				**Cross-Price Elasticities** In response to a price increase by:			
	Brand A	Brand B	Brand C	Brand D	Brand A	Brand B	Brand C	Brand D
Brand A	**2.88%**	**-0.18%**	**-0.43%**	**-0.48%**	**2.63%**	**-0.60%**	**-0.68%**	**-0.83%**
	(3.32%)	(3.53%)	(3.45%)	(3.51%)	(3.97%)	(4.17%)	(4.13%)	(4.12%)
Brand B	**-0.55%**	**2.79%**	**-0.43%**	**-0.48%**	**-0.62%**	**2.55%**	**-0.68%**	**-0.83%**
	(3.45%)	(3.32%)	(3.45%)	(3.51%)	(4.11%)	(3.97%)	(4.13%)	(4.12%)
Brand C	**-0.55%**	**-0.18%**	**2.88%**	**-0.48%**	**-0.62%**	**-0.60%**	**2.63%**	**-0.83%**
	(3.45%)	(3.53%)	(3.32%)	(3.51%)	(4.11%)	(4.17%)	(3.97%)	(4.12%)
Brand D	**-0.55%**	**-0.18%**	**-0.43%**	**2.77%**	**-0.62%**	**-0.60%**	**-0.68%**	**2.53%**
	(3.45%)	(3.53%)	(3.45%)	(3.32%)	(4.11%)	(4.17%)	(4.13%)	(3.97%)
	Own-Brand Promotional Effects				**Own-Brand Promotional Effects**			
		Display Only	Feature Only	Feature & Display		Display Only	Feature Only	Feature & Display
Brand A		**-3.22%**	**-5.23%**	**-6.09%**		**64.74%**	**-2.65%**	**19.53%**
		(6.25%)	(3.38%)	(3.22%)		(110.31%)	(4.68%)	(28.39%)
Brand B		**-3.07%**	**-4.05%**	**-5.13%**		**71.06%**	**-1.77%**	**18.28%**
		(6.53%)	(3.36%)	(3.26%)		(121.74%)	(4.74%)	(26.09%)
Brand C		**-4.18%**	**-5.65%**	**-6.78%**		**70.77%**	**-2.95%**	**20.73%**
		(7.01%)	(3.53%)	(3.47%)		(122.26%)	(4.88%)	(31.49%)
Brand D		**-4.30%**	**-3.73%**	**-4.73%**		**73.70%**	**-1.53%**	**18.45%**
		(7.03%)	(3.21%)	(3.25%)		(126.93%)	(4.54%)	(26.33%)

Notes: The table reports the average percent difference between the true and estimated values across the 5,000 Monte Carlo simulations. The standard deviation of the percent difference is reported in parentheses.