

YUXIN CHEN and SHA YANG*

In this article, the authors propose a Bayesian method for estimating disaggregate choice models using aggregate data. Compared with existing methods, the advantage of the proposed method is that it allows for the analysis of microlevel consumer dynamic behavior, such as the impact of purchase history on current brand choice, when only aggregate-level data are available. The essence of this approach is to simulate latent choice data that are consistent with the observed aggregate data. When the augmented choice data are made available in each iteration of the Markov chain Monte Carlo algorithm, the dynamics of consumer buying behavior can be explicitly modeled. The authors first demonstrate the validity of the method with a series of simulations and then apply the method to an actual store-level data set of consumer purchases of refrigerated orange juice. The authors find a significant amount of dynamics in consumer buying behavior. The proposed method is useful for managers to understand better the consumer purchase dynamics and brand price competition when they have access to aggregate data only.

Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice

Understanding consumer purchase behavior and response to price and promotion is important to marketing managers. The past two decades have witnessed numerous applications of demand models that employ household panel data and aggregate sales data to fulfill this goal. In comparison with aggregate data, household panel data facilitate empirical analysis of individual consumer choice behavior. For example, household panel data provide rich information on purchase history that has been shown to affect consumer current choice decisions significantly (Erdem and Keane

1996; Feinberg, Kahn, and McAlister 1992; Guadagni and Little 1983).

Although household panel data are desirable, they are often available only for a small number of product categories, stores, regions, and weeks, which limits the decision maker's ability to obtain relevant information (Bucklin, Russell, and Srinivasan 1998; Russell and Kamakura 1994). For example, ACNielsen Homescan consists of only 61,500 households in the United States.¹ With the existence of a large number of stores, the amount of household data available at each individual store are extremely limited. Therefore, aggregate data are used in many situations because they are more readily available and are often less costly than household panel data.

Despite the lack of household panel data, store managers may still want to understand microlevel consumer behavior, such as purchase dynamics (i.e., the impact of purchase history on current buying decision), based on store-level aggregate information. This is challenging in the following two ways: First, a sales response model may not address this problem adequately, because it does not have a microlevel underpinning of modeling individual consumer

*Yuxin Chen is Associate Professor of Marketing (e-mail: ychen@stern.nyu.edu), and Sha Yang is Assistant Professor of Marketing (e-mail: shayang@stern.nyu.edu), Leonard N. Stern School of Business, New York University. The authors thank the three anonymous *JMR* reviewers for their constructive comments and suggestions. They also thank Greg Allenby, Eric Bradlow, Pradeep Chintagunta, Andres Musalem, Peter Rossi, K. Sudhir, and seminar participants at Buck Weaver Award Conference at the Massachusetts Institute of Technology, University of Chicago, University of Texas at Austin, University of Michigan, Ohio State University, and the Marketing Science Conference 2004 at Rotterdam for their helpful comments. The authors appear in alphabetical order and contributed equally. Gary J. Russell served as guest editor for this article.

To read and contribute to reader and author dialogue on *JMR*, visit <http://www.marketingpower.com/jmrblog>.

¹See <http://www.acnielsen.com/products/reports/homescan/panelviews/> (accessed on December 2, 2003).

behavior patterns. Second, it is computationally demanding to recover the purchase dynamics because consumer purchase history information is not directly observable from the aggregate information. Because of this difficulty, existing models ignore the purchase dynamics when analyzing aggregate data, though they are built on the consumer utility maximization framework and are able to capture consumer heterogeneity.

In this article, we develop a Bayesian method for modeling disaggregate consumer behavior with aggregate data. The proposed method is especially useful to firms that need a better understanding of microlevel consumer purchase dynamics and accurate estimates of brand price competition at the store level when high-quality panel data are not available. Our approach retains the benefits of discrete choice models, which provide a structural and parsimonious way to model consumer purchase behavior and ensure that the estimates of own- and cross-elasticities (Kamukura and Russell 1989; Rossi and Allenby 1993) are of the correct sign. As do all Bayesian methods, the proposed method facilitates exact, finite-sample inference and does not rely on asymptotic results (Bajari 2003; Berry 2003; Rossi and Allenby 2003).

A key advantage of our method is the ability to model the impact of purchase history on consumer choice decisions with aggregate data. This is achieved through data augmentation (Tanner and Wong 1987), in which we simulate a set of latent choices that are consistent with the observed aggregate data. Because individual choices are not made available through existing estimation methods without using data augmentation, this effect will be difficult to measure.

To demonstrate the validity of the proposed method, we first conduct a series of simulations. In the simulation exercises, we generate the aggregate brand shares from individual consumers' choices, given the prespecified response parameters and distribution of heterogeneity. Using the proposed method, we can recover the true model parameters. We then apply the method to an empirical application based on store-level data of consumer purchases of refrigerated orange juice. We find a significant impact of last-period purchase on current-period purchase.

We organize the remainder of the article as follows: In the next section, we introduce a Bayesian method for estimating disaggregate choice models based on aggregate data, which enables us to capture the consumer purchase dynamics. We then test the validity of the method in a series of simulations and provide an empirical application based on store-level refrigerated orange juice data. The final section offers some concluding remarks.

USING AGGREGATE DATA TO ESTIMATE DISAGGREGATE MODELS

We begin by modeling the data-generating process of the observed aggregate shares. We assume that the observed aggregate share, S_{jt} , of brand j ($j = 0, \dots, J$) at time t ($t = 1, \dots, T$) is generated by M clusters of consumers in the market. Each cluster contains an equal number of consumers, and we assume that consumers within the same cluster have the same utility. Because no disaggregate-level information is observed, to add flexibility to our model setup, we treat a cluster as the minimum unit of analysis at the disaggregate level. As with all previous approaches for estimating dis-

crete choice models with aggregate data, we assume that each consumer purchases, at most, one unit of a product for each period. The utility of brand j at time t for consumers in cluster i ($i = 1, \dots, M$) can be specified as

$$(1) \quad u_{ijt} = \theta_i' x_{ijt} + \varepsilon_{ijt},$$

where x_{ijt} is a vector that includes marketing variables, such as price, feature, and display (it can also include variables on consumer purchase history to capture the purchase dynamics); θ_i is a vector of corresponding coefficients following a multivariate normal distribution with a mean vector $\bar{\theta}$ and a diagonal variance-covariance matrix D (i.e., $\theta_i \sim \text{MVN}[\bar{\theta}, D]$); and ε_{ijt} is the unobserved error term following Type I extreme value distribution. The utility specification in Equation 1 leads to a discrete choice model with a standard logit choice probability,

$$(2) \quad \Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\theta_i' x_{ijt})}{\sum_{k=0}^J \exp(\theta_i' x_{ikt})},$$

where $y_{ijt} = 1$ if consumers in cluster i choose brand j at time t and $y_{ijt} = 0$ if otherwise.

Two general approaches are proposed in the literature to estimate discrete choice models using aggregate data. The first approach is not likelihood based. In this approach, model parameters are estimated by either minimizing the discrepancy between observed market share and predicted market share (Boyd and Mellman 1980; Cardell and Dunbar 1980; Tardiff 1980) or equating the two by introducing unobserved product characteristics into the utility function (Berry 1994; Berry, Levinsohn, and Pakes 1995).

The second approach is to specify a likelihood function of aggregate data (Bodapati and Gupta 2004; Kim 1995; Zenor and Srivastava 1993). According to this approach, the likelihood function based on Equation 1 can be written as

$$(3) \quad L = \prod_{t=1}^T \left\{ \binom{M}{O_{0t}, \dots, O_{Jt}} \prod_{j=0}^J \left[\int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i \right]^{O_{jt}} \right\},$$

where

$$\binom{M}{O_{0t}, \dots, O_{Jt}}$$

is the multinomial coefficient, $f(\theta_i | \bar{\theta}, D)$ is the probability density function of θ_i , and O_{jt} is the number of clusters among M that choose brand j at time t . The term O_{jt} can be operationalized as $O_{jt} = \text{int}(S_{jt}M + .5)$, for $j = 0, \dots, J - 1$, and $O_{Jt} = M - \sum_{j=0}^{J-1} O_{jt}$. This likelihood function is based on the view that the M clusters are exchangeable, because we observe only aggregate data and each cluster has an expected brand choice probability $\int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i$.

Although the estimation of the random utility model with aggregate data is made feasible by the two existing approaches, it is not straightforward to incorporate the consumer purchase history into the underlying discrete choice model, because such information is not directly available. To overcome this difficulty, we propose a hierarchical Bayesian model to treat individual choices as latent variables and to augment them from the observed aggregate

information. Such an approach will facilitate the study of microlevel consumer purchase dynamics using store-level share data.

The basic intuition of our approach is to represent the M clusters of consumers with a panel of R clusters whose choices are augmented. The average choice probability of this panel approximates the expected choice probability of the remaining $M-R$ clusters. Although the $M-R$ clusters remain exchangeable, the R clusters are no longer exchangeable because their choices are augmented. This feature facilitates the modeling of purchase dynamics with aggregate data because the augmented purchase history now enters the utility specification directly when calculating the choice probability s_{ijt} .

Formally, our approach uses the likelihood specification in Equation 3 as the starting point but departs from it by assuming that there are R representative clusters among M whose choices, which are denoted as y_{rjt} , where $r \in R$ and $R \leq M$, are to be augmented. Given the observed aggregate data, a permissible purchase history, h , on R is a set of y_{rjt} ($r = 1, \dots, R, j = 0, \dots, J$, and $t = 1, \dots, T$) such that the condition

$$(4) \quad \sum_{r=1}^R y_{rjt} \leq O_{jt}$$

holds for all $j = 0, \dots, J$ and $t = 1, \dots, T$. The condition imposed in Equation 4 is the result of R being a subset of M . In our proposed estimation approach, y_{rjt} and, consequently, h are augmented. Let H be the collection of all h . We can write the likelihood of the observed aggregate share data as

$$(5) \quad L = \sum_{h \in H} (L_{R|h} \times L_{M-R|h}).$$

In Equation 5, $L_{R|h}$ is the likelihood of a permissible purchase history, h , of the panel of R representative clusters,

$$(6) \quad L_{R|h} = \prod_{r=1}^R \int \left[\prod_{t=1}^T \prod_{j=0}^J (s_{rjt})^{y_{rjt}} f(\theta_r | \bar{\theta}, D) \right] d\theta_r.$$

Note that the likelihood specified in Equation 6 implies that the R representative clusters are no longer treated as exchangeable. This is because we augment the purchase history of the R clusters. The remaining $M-R$ clusters are still viewed as exchangeable because their purchase history is not augmented. Therefore, analogous to Equation 3,

$$(7) \quad L_{M-R|h} = \prod_{t=1}^T \left\{ \binom{M-R}{Z_{0t}, \dots, Z_{Jt}} \prod_{j=0}^J \left[\int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i \right]^{Z_{jt}} \right\},$$

where

$$Z_{jt} = O_{jt} - \sum_{r=1}^R y_{rjt}, \text{ and } \binom{M-R}{Z_{0t}, \dots, Z_{Jt}}$$

is the multinomial coefficient.

Because the R clusters are assumed to be representative, we can approximate the brand choice probability in Equation 7, $\int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i$, with the average brand choice probability of the R clusters; that is,

$$(8) \quad \int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i = \sum_{r=1}^R s_{rjt} / R.$$

With $L_{R|h}$ and $L_{M-R|h}$, the likelihood of purchases made by the M clusters given a permissible purchase history, h , on the R clusters is $L_{R|h} \times L_{M-R|h}$. Summing over all h gives the likelihood on the aggregate data, as we show in Equation 5.

A few comments on our model setup are in order. First, a key feature of the proposed approach is that the choices of the panel of R clusters are augmented. This facilitates the modeling of consumer purchase dynamics with aggregate data because the augmented purchase history can now enter the utility specification directly to calculate s_{ijt} on the right-hand side of Equation 8. Without augmentation of the choices made by R clusters, when s_{ijt} is a function of purchase history, it would be necessary to integrate over all possible paths of purchase history to evaluate $\int s_{ijt} f(\theta_i | \bar{\theta}, D) d\theta_i$ on the left-hand side of Equation 8, which is impractical.

Second, we want to emphasize that the heterogeneity distribution in our model is assumed over M clusters and applies to both the R and the $M-R$ clusters. That is, we assume that the multinomial choice probability of the $M-R$ clusters resulting from integration over the heterogeneity distribution can be approximated by the average choice probability from the augmented panel of R clusters. Essentially, the heterogeneity distribution over R clusters can represent the heterogeneity distribution over M clusters.

Third, we can let $R = M$ in principle. In this case, the likelihood in Equation 5 becomes the same as the likelihood function of a discrete choice model with panel data. However, the drawback of setting $R = M$ is that Equation 4 is always binding. Consequently, the algorithm of augmenting y_{rjt} becomes inefficient because the exact constraint implied by Equation 4 must apply.² If we let $R < M$, Equation 4 does not need to be binding. As a result, the estimation algorithm becomes efficient because we can now use the standard logit probability to generate candidates of y_{rjt} with a high acceptance rate.

Finally, the choices of M and R are important to our model. Theoretically speaking, a large R relative to M reduces the acceptance rate in generating augmented choices and makes the algorithm less efficient. In contrast, a small R relative to M makes the algorithm more efficient, but such an R may not represent M well. In practice, our simulation results suggest that when M is close to its true value, the results are not sensitive to the choice of R . However, if M is far from its true value, the algorithm fails to converge for some values of R within a reasonable number of iterations. Therefore, in an empirical application in which the value of M is unknown, if the estimation results are robust when varying R for a chosen M , it suggests that

²In a previous version of this article, we developed the estimation algorithm under $R = M$, which is less efficient than the one presented herein.

the chosen M is reasonable. In the next section, we provide more guidance on the choice of M when it is unknown.

We now discuss how to estimate the model as given in Equations 5–8. The likelihood specification in Equation 5 is cumbersome because of the high dimensions of integration due to the consumer heterogeneity distribution of θ_i and, more important, the large number of combinations of permissible purchase history (h) that are consistent with the aggregate shares. Because a direct evaluation of the likelihood is costly, we proceed with Bayesian analysis using data augmentation (Albert and Chib 1993; Tanner and Wong 1987). In other words, rather than integrating out individual choices and response parameters, we treat them as any other unobserved model parameters and use them as conditioning arguments in hierarchical Bayesian analysis. This leads to a substantial simplification in estimation.

Our Bayesian analysis of the observed aggregate data proceeds by specifying the joint distribution of all model parameters. Estimation is carried out by setting up a Markov chain and iteratively sampling from the conditional distributions of model parameters. The joint posterior distribution can be written as

$$(9) \quad f(\{y_t\}, \{\theta_r\}, \delta\{S_t\}, \{x_{rt}\}) \propto \prod_{t=1}^T \{f(S_t|y_t, \theta_1, \dots, \theta_R) \prod_{r=1}^R [f(y_{rt}|\theta_r, x_{rt})f(\theta_r|\bar{\theta}, D)]f(\bar{\theta}, D)\},$$

where y_t is the vector of y_{ijt} , $f(S_t|y_t, \theta_1, \dots, \theta_R) = L_{M-R|h,t}$, $f(y_{rt}|\theta_r, x_{rt}) = \prod_{j=0}^J (s_{rjt})^{y_{rjt}}$ is the choice probability at time t of cluster r whose choices are augmented, $f(\theta_r|\bar{\theta}, D)$ is the heterogeneity distribution, and $f(\bar{\theta}, D)$ is the prior distribution. The Markov chain involves a sequence of draws from the full conditional distributions of the model (see Gelfand and Smith 1990). We describe a Markov chain Monte Carlo (MCMC) algorithm for generating draws of the model parameters next (for details, see the Appendix).

The key to our algorithm is to augment individual choices by generating y_t conditional on other parameters in the model. From Equation 9, the conditional distribution of y_t is proportional to $f(S_t|y_t, \theta_1, \dots, \theta_R) \prod_{r=1}^R f(y_{rt}|\theta_r, x_{rt})$. Because $f(y_{rt}|\theta_r, x_{rt}) = \prod_{j=0}^J (s_{rjt})^{y_{rjt}}$, we can generate the candidate draws of y_{rt} ($r = 1, \dots, R$) from a discrete distribution with $J + 1$ outcomes. Each outcome is a vector with only one element being 1 to indicate the chosen alternative; the rest are 0. The probability of each outcome is s_{rjt} ($j = 0, \dots, J$), which is the logit probability. The candidate draws of y_{rt} are qualified if the resultant Z_{jt} is nonnegative for all j (i.e., Equation 4 holds). Otherwise, another set of draws of y_{rt} is generated. After a qualified draw of y_t is obtained, the acceptance probability of this candidate draw is given by the ratio of $f(S_t|y_t, \theta_1, \dots, \theta_R)$ calculated from the current draw and the one calculated from the y_t of the previous iteration. This is the third family of Metropolis–Hasting algorithms, as described in the work of Chib and Greenberg (1995, p. 330).

After y_t is generated, θ_r can be generated using a random walk chain Metropolis–Hastings algorithm. Routines for generating the hyperparameters ($\bar{\theta}, D$) are similar to that

which Allenby and Rossi (1999) describe for standard random coefficients models. In the next section, we test the proposed method in simulations and discuss the implementation issues.

SIMULATION STUDIES

To test whether the proposed method can recover the true model parameters in different situations, we simulated aggregate shares on the basis of individual consumer choices that are generated from prespecified parameters. We conducted simulations on models with and without purchase dynamics. Given the space limitations, we report only the results from the model with purchase dynamics. The random coefficient logit model without purchase dynamics is widely used in the empirical industrial organization research on aggregate data. Readers who are interested in such a model can access the Web Appendix (see <http://www.marketingpower.com/content84060.php>) for the corresponding simulation results.

In the simulation study with purchase dynamics, we include two brands and a no-purchase option. The no-purchase option is denoted as $j = 0$, and the associated utility is specified as³

$$(10) \quad u_{i0t} = \varepsilon_{i0t}.$$

The simulation is carried out in the following context: We simulate share data for two brands and a no-purchase option in 100 periods aggregated from choices based on 5000 θ_i 's. The vector x is composed of two brand intercepts (x_1 and x_2), the price (x_3) generated from a normal distribution with mean equal to 0 and standard deviation equal to 1, and the dynamic covariate denoted as LAST (x_4), which is measured on the basis of whether there is a purchase of any of the brands in the last period ($x_4 = 1$ if there was a purchase of any of the brands in the last period, and $x_4 = 0$ if otherwise). For convenience, $x_4 = 0$ for the first period across all consumers. The true mean preference parameter $\bar{\theta}$ and heterogeneity matrix D take the following values:

$$(11) \quad \bar{\theta} = (1, 1, -1, .5), \text{ and}$$

$$(12) \quad D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & .5 \end{bmatrix}.$$

Choices (y_{ijt}) are simulated with the latent brand utility, as we specify in Equation 1. Aggregate brand shares are formed by

$$(13) \quad S_{jt} = \sum_{i=1}^{5000} y_{ijt} / 5000 \text{ (for all } j \text{ and } t).$$

The estimation task is to estimate $\bar{\theta}$ and D on the basis of the simulated share S_{jt} . If the method is valid, we should get

³An alternative way to model the no-purchase option is through a nested logit setup. This results in a different model structure from the one we present here. Our simulation results show that the proposed method also applies to the nested logit model. We include the results in the Web Appendix (see <http://www.marketingpower.com/content84060.php>).

the estimated parameter values close to the true ones set in the simulation.

The estimation is based on $M = 5000$ and $R = 100$. We ran the Markov chains for 50,000 iterations and saved every tenth draw for the parameters of interest. We set the starting value for θ as $(0, 0, 0, 0)$ and for the four diagonal elements in D as .1. The starting values for y_{jt} are randomly generated with an equal choice probability of the three alternatives at the individual level, given the consistency condition ($\sum_{t=1}^R y_{jkt} \leq O_{jt}$ for all j and t). The last 25,000 draws are used to calculate the posterior means and standard deviations of the parameters.

Table 1 reports the results. The table shows that the true values of all parameters lie within the 95% confidence intervals of their posterior mean estimates. Our additional simulations also show that the results are not sensitive to the choice of starting values. These results provide evidence for the validity of the proposed method.

We examined the sensitivity of our estimation results with respect to the two control variables in the experiment, T (the length of the purchase history) and R .⁴ We find that the true values are all covered in the 95% confidence interval of their posterior mean estimates. As we expected, when T or R increases, the posterior standard deviations of the parameter estimates tend to decrease gradually. The proposed method achieves reasonable accuracy even with small T ($T = 50$) and R ($R = 50$).

We now provide some discussion and cautionary notes on the model and simulation exercises. All the results we obtained here assume that M is known. However, this is often not true in an empirical application. Therefore, it is important to test the sensitivity of the estimation results to M . For the simulated data generated from 5000 θ_i 's, we can recover the true values of all the parameters when M is reasonably close to its true value. However, the MCMC fails to converge within a reasonable number of iterations if we choose an M that is far different from the true value of 5000. For example, this happens when $M = 50,000$ or $M = 200$. In general, our simulation studies for models either with or without purchase dynamics suggest that when M is reasonably close to its true value, the results are not sensi-

tive to the choice of R . However, if M is far from its true value, the MCMC fails to converge within a reasonable number of runs for some values of R . This approach offers some guidelines for applying the model in an empirical application in which M is often unknown. Under such a circumstance, we suggest a two-step procedure to select a reasonable M . The first step is to find an M under which the MCMC converges. We can do so by starting from an M that is close to a reasonable value of R (e.g., $R = 100$) and increasing the value of M subsequently. The second step is to vary R to check whether the estimation results are sensitive to the choice of R , given the M chosen in the first step. If the results are not sensitive to R , the chosen M is reasonable. Otherwise, we need to go back to the first step and find a new candidate of M .

Next, we acknowledge some limitations in our model setup. First, we assume a diagonal heterogeneity covariance matrix that is not as general as a full covariance matrix. Our simulation results indicate that the accuracy of the off-diagonal elements in a full covariance heterogeneity matrix cannot be fully guaranteed. It is possibly because of this identification difficulty that all the previous studies have also adopted the assumption of a diagonal heterogeneity covariance matrix.

Second, the model we propose assumes that consumer heterogeneity follows a normal distribution as in a standard random-coefficients framework. A concern is that if the market is composed of distinct consumer segments (e.g., heterogeneity being specified as a mixture of widely separated multivariate normal distributions), distortion will occur when we proceed with the unimodal assumption on the heterogeneity distribution in estimation. We conduct a simulation by assuming a unimodal distribution for consumer heterogeneity with the true distribution being a mixture of two widely separated normal distributions. Our results indicate that estimates for the preference parameters all converge to the average parameter values of the two segments. However, there is a significance bias for the heterogeneity distribution. This is consistent with previous findings with panel data (Allenby, Arora, and Ginter 1998). Misspecification of the heterogeneity distribution can lead to biased inferences. Therefore, researchers should take caution when using the proposed method.

⁴Simulation results related to the discussion in the remainder of this section are in the Web Appendix (see <http://www.marketingpower.com/content84060.php>).

Table 1
SIMULATION WITH PURCHASE DYNAMICS: POSTERIOR MEANS AND STANDARD DEVIATIONS FOR THE ESTIMATES OF θ AND D

	<i>True Values</i>	<i>Estimates</i>
$\bar{\theta}_1$ (Brand 1)	1.0	1.06 (.12)
$\bar{\theta}_2$ (Brand 2)	1.0	1.06 (.12)
$\bar{\theta}_3$ (Price)	-1.0	-1.03 (.11)
$\bar{\theta}_4$ (LAST)	.5	.48 (.11)
$D_{1,1}$	1.0	1.02 (.30)
$D_{2,2}$	1.0	1.14 (.37)
$D_{3,3}$	1.0	1.06 (.19)
$D_{4,4}$.5	.64 (.16)

AN EMPIRICAL APPLICATION

In this section, we illustrate our model and estimation approach with an empirical application that achieves the following three objectives: First, we further demonstrate the ability of this method in estimating the impact of purchase history on consumer brand choices. Specifically, we allow consumer brand choices to be a function of the last-period purchase information. Second, we estimate the brand price competition structure at the store level, using the proposed data augmentation method. As we discussed previously, firms often have access to high-quality store-level data but limited household-level information. Therefore, the proposed method will be attractive to firms in understanding brand competition and setting optimal category prices at the store level. Third, we compare two alternative approaches of modeling the no-purchase option. The first approach is to treat the no-purchase option as another choice alternative

and to set its mean utility to zero (Chintagunta 2002; Sudhir 2001). The second approach is to adopt a nested logit specification (Bucklin and Lattin 1991). In this approach, we model the purchase/no-purchase decision as a binary choice with the following purchase probability:

$$(14) \quad pr_{it}(\text{purchase}) = \frac{\exp(\gamma_{i0} + \gamma_{i1}CV_{it} + \gamma_{i2}LAST_{it})}{1 + \exp(\gamma_{i0} + \gamma_{i1}CV_{it} + \gamma_{i2}LAST_{it})}$$

where CV_{it} is the category value, which takes the form of $\ln[\sum_j \exp(V_{ijt})]$, and $LAST$ refers to whether there is a purchase of any of the brands in the last period. For identification purposes, we fix γ_{i0} to be zero.

We use store-level data from the Dominick's Finer Foods chains based in the Chicago market to demonstrate the proposed Bayesian method of estimation. The Dominick's Finer Foods database contains detailed information on Universal Product Code-level (UPC-level) and store-level sales volume, price, and promotion activity on a weekly basis. We analyze the refrigerated orange juice category in one store. Our data span from June 1991 to May 1993, for a total of 104 weeks. Descriptive statistics appear in Table 2. Note that price and promotion frequency for each brand are calculated as weighted averages across different UPCs under that brand by UPC-level sales volume. To calculate the outside good share, we first obtain the total inside good share (based on volume measured in ounces) and then subtract it from 1. The inside good share can be estimated by dividing the actual weekly sales volume by the potential weekly sales volume. The potential weekly sales volume is calculated by multiplying the average weekly sales by the average buying cycle (approximately 3.6 weeks in the Chicago area, according to ACNielsen Homescan Consumer*Facts).⁵

We focus our analysis on the top three brands: Minute Maid, Dominick's store brand, and Tropicana. The data reveal that the store brand achieves a higher market share than both national brands, has a lower unit price, and has a moderate deal frequency. Comparing the two national brands, we observe that Tropicana has a premium price, engages in more promotion activities, and achieves a higher market share than Minute Maid.

We calibrate three models. We ran the Markov chain for 500,000 iterations ($M = 200$, $R = 100$).⁶ The last 250,000

⁵Given the assumption that each consumer buys, at most, one unit of the product, the potential weekly sales volume is equal to the number of consumers in the market. Because the average consumer purchase cycle is 3.6 weeks, this implies that 1/3.6 proportion of consumers in the market is expected to purchase in a given week.

⁶When M was assumed to be 200, we found similar results for different values of R . We also obtained similar results for $M \leq 1000$. For larger M ,

Table 2

SAMPLE STATISTICS FOR REFRIGERATED ORANGE JUICE DATA (SAMPLE MEANS AND STANDARD DEVIATIONS)

Brands	Choice Share	Average Price (cents/ounce)	Promotion Frequency
Minute Maid	.07 (.08)	3.35 (.40)	.33 (.37)
Dominick's	.11 (.07)	2.39 (.28)	.38 (.37)
Tropicana	.10 (.07)	3.75 (.42)	.41 (.33)
Outside good	.72 (.10)	N.A.	N.A.

Notes: N.A. = not applicable.

iterations were saved to compute the estimation results. Specifically, Model 1 is a standard random coefficients logit model, with the mean utility of the no-purchase option set to zero. Model 2 adopts the same structure as Model 1 but adds an additional explanatory variable to capture the effect of last purchase information (measured as whether there is a purchase of the three brands in the last period). Model 3 jointly models the purchase/no-purchase decision and the brand choice decision conditional on purchase, as in a nested logit model. The utility function in Model 3 is specified to be the same as that in Model 1. The purchase/no-purchase decision is modeled as a binary choice with the purchase probability specified in Equation 14. Fit statistics indicate the following:⁷ We find that the in-sample model fit of Model 1 (-9260.75) is smaller than both Model 2 (-8407.33) and Model 3 (-8628.40), suggesting the importance of capturing the effect of last-period purchase. In addition, the higher fit statistics for Model 2 than Model 3 indicate that it is better to model the no-choice option simply as another choice alternative for this data set.

Estimates of the preference parameters for all three models appear in Table 3. Because Model 2 is the best-fitting one, our discussion is based on the results from that model. Overall, the estimates have strong face validity. The two national brands have larger intercept estimates than the store brand. For the two national brands, Tropicana has a

the MCMC failed to converge for some values of R . From our discussion in the simulation study, the choice of $M = 200$ is reasonable.

⁷We measured model fit as the log-marginal density, which we calculated using Newton and Raftery's (1994, p. 21) importance sampling method.

Table 3
POSTERIOR MEANS (POSTERIOR STANDARD DEVIATIONS)
OF PREFERENCE PARAMETER ESTIMATES FOR
REFRIGERATED ORANGE JUICE DATA

Coefficients for	Model 1 (Logit Model Without Dynamics)	Model 2 (Logit Model with Dynamics)	Model 3 (Nested Logit Model with Dynamics)
<i>Brand Choice Model</i>			
Minute Maid	.92 (.16)	1.15 (.19)	.83 (.28)
Dominick's	.31 (.15)	.44 (.15)	-.04 (.24)
Tropicana	1.63 (.18)	1.83 (.20)	1.61 (.29)
Price	-1.19 (.06)	-1.29 (.07)	-1.49 (.08)
Promotion	.69 (.15)	.69 (.11)	.85 (.19)
Last	N.A.	-1.54 (.43)	N.A.
<i>Purchase/No-Purchase Model</i>			
Category value	N.A.	N.A.	.54 (.04)
Last	N.A.	N.A.	-1.64 (.49)

Notes: $CV = \ln[\sum_j \exp(V_j)]$. Last = 1 if the consumer bought one of the three focal brands last week, and Last = 0 if otherwise. N.A. = not applicable.

larger intercept estimate. The price and promotion coefficient estimates are significant and have the expected signs. A notable finding is the effect of last purchase. As we discussed previously, the ability to model such an effect is an advantage of the proposed method. The coefficient estimate for the last purchase is significantly negative, indicating that a consumer is less likely to purchase from the three brands again if he or she purchased any of them the previous week.

Table 4 reports the heterogeneity estimates. We find significant estimates for the consumer heterogeneity of all parameters across the three models. There is also a significant difference in the heterogeneity estimate of promotion sensitivity across the three models. In addition, we find a significant difference in the heterogeneity estimates in the model that does not account for consumer purchase dynamics (Model 1) compared with the model that captures the impact of purchase history and achieves the best fit (Model 2). Notably, the heterogeneity estimate of the last-purchase coefficient is large. This may reflect a large difference in consumption patterns among consumers. For example, some households may consume refrigerated orange juice more frequently and purchase every week, and others may consume and purchase less frequently. Such variation can be captured in the heterogeneity of the last-purchase coefficient.

Price elasticity provides important information in helping firms better understand market competition and design optimal prices. We report the estimated price elasticity matrices in Table 5. The elasticity matrix from Model 2 displays the following patterns: First, the two national brands have larger own-price elasticities than the store brand. Second, the price change of Tropicana affects the shares of the other

Table 5
PRICE ELASTICITY MATRIX FOR REFRIGERATED ORANGE JUICE DATA

Products	Effect of Minute Maid's Price on ...	Effect of Dominick's Price on ...	Effect of Tropicana's Price on ...
Minute Maid	-4.40	.42	.62
Dominick's	.39	-3.18	.59
Tropicana	.40	.43	-4.68

two brands more than the other way around. The impact of a price change of Tropicana on the share of Minute Maid is .62 (Minute Maid's effect on Tropicana is .40). The impact of a price change of Tropicana on the share of Dominick's store brand is .59 (the store brand's effect on Tropicana is .43). This finding is consistent with the asymmetric price competition pattern that Blattberg and Wisniewski (1989) document.

SUMMARY AND CONCLUDING REMARKS

In this article, we develop a Bayesian method for estimating disaggregate choice models using aggregate data. Our approach takes advantage of the low cost and easy accessibility of aggregate data and enjoys the desirable features of discrete choice modeling with household scanner panel data, such as being parsimonious, structural, and often devoid of the wrong-sign problem. Compared with other existing methods, the unique advantage of the proposed method is that it allows for the analysis of microlevel consumer behavior, such as the impact of purchase history, when only aggregate-level data are available. Our simulation experiments and empirical application establish the validity of the proposed estimation approach under our model specification. This method is especially attractive to managers in better understanding microlevel consumer purchase dynamics and store-level brand competition when household panel data are not available.

There are a few caveats of the model and the proposed estimation approach that we want to address. As with previous methods for estimating choice models with aggregate data, we cannot identify well the off-diagonal elements in a full heterogeneity covariance matrix. Our simulation results also indicate that accurate estimates on the heterogeneity distribution estimates rely on the unimodal assumption, violation of which leads to biased inferences on the heterogeneity though not on the grand mean of the preference parameters. Finally, the implementation of our proposed estimation approach requires choosing a value of M, and therefore some sensitivity analysis on M is necessary.

The proposed method opens up new opportunities for researchers interested in estimating microlevel models with aggregate data. First, various purchase history-related effects, such as consumer loyalty (Guadagni and Little 1983) and variety seeking (Feinberg, Kahn, and McAlister 1992), can potentially be incorporated into the microlevel model specification and estimated with aggregate data using the proposed framework. Second, the method can be extended to combining aggregate and disaggregate data sources (Berry, Levinsohn, and Pakes 2004; Bucklin, Russell, and Srinivasan 1998; Chintagunta and Dubé 2005; Russell and Kamakura 1994; Swait and Andrews 2003)

Table 4

POSTERIOR MEANS (POSTERIOR STANDARD DEVIATIONS) OF HETEROGENEITY PARAMETER ESTIMATES FOR REFRIGERATED ORANGE JUICE DATA

Coefficients for	Model 1 (Logit Model Without Dynamics)	Model 2 (Logit Model with Dynamics)	Model 3 (Nested Logit Model with Dynamics)
<i>Brand Choice Model</i>			
Minute Maid	.48 (.13)	.35 (.11)	.41 (.14)
Dominick's	.69 (.22)	.42 (.13)	.63 (.22)
Tropicana	.54 (.15)	.49 (.18)	.60 (.22)
Price	.18 (.03)	.14 (.03)	.22 (.05)
Promotion	1.71 (.42)	.80 (.23)	2.70 (.77)
Last	N.A.	13.51 (2.83)	N.A.
<i>Purchase/No-Purchase Model</i>			
Category value	N.A.	N.A.	.09 (.01)
Last	N.A.	N.A.	8.40 (2.03)

Notes: CV = $\ln[\sum_j \exp(V_j)]$. Last = 1 if the consumer bought one of the three focal brands last week, and Last = 0 if otherwise. N.A. = not applicable.

because of the unique advantages of using each of the two types of data. Finally, the method can be applied to estimating simultaneous demand and supply models with aggregate data, coupled with the approach that Yang, Chen, and Allenby (2003) propose for estimating such models with scanner panel data, which can potentially contribute to the empirical industrial organization literature.

APPENDIX: MCMC ESTIMATION

Estimation is carried out by sequentially generating draws from the following distributions:

1. Generate $(y_t, t = 1, \dots, T)$.

We use the Metropolis–Hastings algorithm to generate y_t . The posterior distribution of y_t conditional on all other parameters is proportional to

$$\prod_{\tau=t}^{t'} \left[\binom{M-R}{Z_{0\tau}, \dots, Z_{J\tau}} \prod_{j=0}^J \left(\sum_{r=1}^R s_{rj\tau} / R \right)^{Z_{jt}} \right] \left[\prod_{r=1}^R \prod_{j=0}^J s_{rj\tau}^{y_{rj\tau}} \right]$$

where

$$s_{rjt} = \frac{\exp(\theta_r' x_{jt})}{\sum_k \exp(\theta_r' x_{kt})}$$

and $t + 1, \dots, t'$ are the periods in which consumer utility is affected by choice y_t . If there is no dynamic effect, $t' = t$. If the dynamic effect carries over for only one period, as in the model we specified, $t' = t + 1$.

We generate the candidate draws of $y_{rt}, r = 1, \dots, R$, from a discrete distribution with $J + 1$ outcomes. Each outcome is a vector with only one element being 1 to indicate the chosen alternative; the rest are 0. The probability of each outcome is $s_{rjt} (j = 0, \dots, J)$, which is the logit probability. The candidate draw of y_{rt} qualifies if the resultant Z_{jt} is nonnegative for all j . Otherwise, a new set of $y_{rt}, r = 1, \dots, R$, is generated until the resultant y_{rt} qualifies. Let $y_t^{(p)}$ denote the previous draw; the next draw, $y_t^{(n)}$, is accepted with probability α given by

$$\min \left\{ \prod_{\tau=t}^{t'} \left[\binom{M-R}{Z_{0\tau}, \dots, Z_{J\tau}} \prod_{j=0}^J \left(\sum_{r=1}^R s_{rj\tau} / R \right)^{Z_{jt}} \right] \frac{\prod_{r=1}^R \prod_{j=0}^J s_{rj\tau}^{y_{rj\tau}^{(n)}}}{\prod_{r=1}^R \prod_{j=0}^J s_{rj\tau}^{y_{rj\tau}^{(p)}}} \right] y_t^{(n)}, 1 \right\}$$

2. Generate $(\theta_r, r = 1, \dots, R)$.

The likelihood function is

$$l(\theta_r) \propto \prod_{t=1}^T \left[\prod_{j=0}^J \left(\sum_{r=1}^R s_{rjt} / R \right)^{Z_{jt}} \prod_{j=0}^J s_{rjt}^{y_{rjt}} \right] \text{ and } s_{rjt} = \frac{\exp(\theta_r' x_{jt})}{\sum_k \exp(\theta_r' x_{kt})}$$

and the posterior distribution is

$$f(\theta_r) \propto |D|^{-1/2} \exp[-1/2(\theta_r - \bar{\theta})' D^{-1}(\theta_r - \bar{\theta})] l(\theta_r),$$

assuming that the prior distribution of θ_r is multivariate normal with mean $\bar{\theta}$ and covariance matrix D . We use the Metropolis–Hastings algorithm with a random walk chain to generate draws of θ_r (see Chib and Greenberg 1995, p. 330, Method 1). Let $\theta_r^{(p)}$ denote the previous draw; then, the next draw $\theta_r^{(n)}$ is given by

$$\theta_r^{(n)} = \theta_r^{(p)} + \Delta,$$

with the accepting probability α given by

$$\min \left\{ \frac{\exp[-1/2(\theta_r^{(n)} - \bar{\theta})' D^{-1}(\theta_r^{(n)} - \bar{\theta})] l(\theta_r^{(n)})}{\exp[-1/2(\theta_r^{(p)} - \bar{\theta})' D^{-1}(\theta_r^{(p)} - \bar{\theta})] l(\theta_r^{(p)})}, 1 \right\},$$

and Δ is a draw from the density Normal(0, .015I), where I is the identity matrix.

3. Generate the diagonal elements of $D, D_{k,k} (k = 1, \dots, K)$, where K is the dimension of θ_r .

The posterior distribution of $D_{k,k}$ is inverted gamma

$$f(D_{k,k} | \theta_{rk}, \bar{\theta}_k) \propto \text{Inverted gamma}(a, b),$$

where

$$a = s_0 + R/2 \quad (s_0 = 3) \text{ and } b = \frac{2}{\sum_{r=1}^R (\theta_{rk} - \bar{\theta}_k)^2 + 2/q_0} \quad (q_0 = .2)$$

4. Generate $\bar{\theta}$.

$$f(\bar{\theta} | \theta_r, D) = \text{MVN}(v, \Psi),$$

where

$$v = \Psi (\sum_{r=1}^R \theta_r / R + D_0^{-1} \theta_0), \Psi = (D_0^{-1} + R D^{-1})^{-1}, \theta_0 = (0, 0, \dots, 0)', \text{ and } D_0 = 100I.$$

REFERENCES

Albert, James and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88 (422), 669–79.
 Allenby, Greg M., Neeraj Arora, and James L. Ginter (1998), "On the Heterogeneity of Demand," *Journal of Marketing Research*, 35 (August), 384–89.

- and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89 (1–2), 57–78.
- Bajari, Patrick (2003), "Comment on: Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1 (3), 277–83.
- Berry, Steven (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (2), 242–62.
- (2003), "Comment on: Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1 (3), 285–91.
- , James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–90.
- , ———, and ——— (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112 (1), 68–104.
- Blattberg, Robert C. and Kenneth J. Wisniewski (1989), "Price-Induced Patterns of Competition," *Marketing Science*, 8 (4), 291–309.
- Bodapati, Anand V. and Sachin Gupta (2004), "The Recoverability of Segmentation Structure from Store-Level Aggregate Data," *Journal of Marketing Research*, 41 (August), 351–64.
- Boyd, J. Hayden and Robert E. Mellman (1980), "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis," *Transportation Research*, 14A, 367–78.
- Bucklin, E. Randolph and James M. Lattin (1991), "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, 10 (1), 24–39.
- , Gary J. Russell, and V. Srinivasan (1998), "A Relationship Between Price Elasticities and Brand Switching Probabilities," *Journal of Marketing Research*, 35 (February), 99–113.
- Cardell, N. Scott and Frederick C. Dunbar (1980), "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research*, 14A, 423–34.
- Chib, Siddhartha and Edward Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49 (5), 327–35.
- Chintagunta, Pradeep (2002), "Investigating Category Pricing Behavior in a Retail Chain," *Journal of Marketing Research*, 39 (May), 141–54.
- and Jean-Pierre Dubé (2005), "Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data," *Journal of Marketing Research*, 42 (August), 368–79.
- Erdem, Tulin and Michael P. Keane (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.
- Feinberg, Fred M., Barbara E. Kahn, and Lee McAlister (1992), "Market Share Response When Consumers Seek Variety," *Journal of Marketing Research*, 29 (May), 227–37.
- Gelfand, Alan E. and Adrian F.M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398–409.
- Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203–238.
- Gupta, Sachin, Pradeep Chintagunta, Anil Kaul, and Dick Wittink (1996), "Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data," *Journal of Marketing Research*, 33 (November), 383–98.
- Kamakura, Wagner and Gary Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.
- Kim, Byung-Do (1995), "Incorporating Heterogeneity with Store-Level Aggregate Data," *Marketing Letters*, 6 (2), 159–69.
- Newton, M.A. and A.E. Raftery (1994), "Approximating Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series B*, 56 (1), 3–48.
- Rossi, Peter E. and Greg M. Allenby (1993), "A Bayesian Approach to Estimating Household Parameters," *Journal of Marketing Research*, 30 (May), 171–82.
- and ——— (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304–328.
- Russell, Gary J. and Wagner A. Kamakura (1994), "Understanding Brand Competition with Micro and Macro Scanner Data," *Journal of Marketing Research*, 31 (May), 289–303.
- Sudhir, K. (2001), "Structural Analysis of Competitive Pricing in the Presence of a Strategic Retailer," *Marketing Science*, 20 (3), 244–64.
- Swait, Joffre and Rick Andrews (2003), "Enriching Scanner Panel Models with Choice Experiments," *Marketing Science*, 22 (3), 422–60.
- Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (398), 528–40.
- Tardiff, Timothy J. (1980), "Vehicle Choice Models: Review of Previous Studies and Directions for Further Research," *Transportation Research*, 14A, 327–35.
- Yang, Sha, Yuxin Chen, and Greg M. Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1 (3), 251–75.
- Zenor, Michael J. and Rajendra Srivastava (1993), "Inferring Market Structure with Aggregate Data: A Latent Segment Logit Approach," *Journal of Marketing Research*, 30 (August), 369–79.

Copyright of *Journal of Marketing Research* (JMR) is the property of American Marketing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.