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

---

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 a better understanding of microlevel consumer purchase history into the underlying discrete choice model with a standard logit choice probability,

$$\Pr(y_{ijt} = 1) = s_{ijt} = \frac{\exp(\theta_{ijt}^\prime x_{ijt})}{\sum_k \exp(\theta_{ik}^\prime x_{ikt})},$$

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

$$L = \prod_{t=1}^{T} \left[ \prod_{j=0}^{J} \prod_{h=0}^{H} \int s_{ijt} f(\theta_{ijt}, D) d\theta_{ijt} \right]^{O_{ijt}}$$

where

$$\frac{\theta_{ijt}}{\theta_{ijt}}$$

is the multinomial coefficient, $f(\theta_{ijt}, D)$ is the probability density function of $\theta_{ijt}$ and $O_{ijt}$ is the number of clusters among $M$ that choose brand $j$ at time $t$. The term $O_{ijt}$ can be operationalized as $O_{ijt} = \int s_{ijt} f(\theta_{ijt}, D) d\theta_{ijt}$.

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 \( y_{rjt} \) (\( r = 1, \ldots, R \), \( j = 0, \ldots, J \), and \( t = 1, \ldots, T \)) such that the condition

\[
\sum_{r=1}^{R} y_{rjt} \leq O_{jt}
\]

holds for all \( j = 0, \ldots, J \) and \( t = 1, \ldots, T \). 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

\[
L = \sum_{h \in H} (L_{R|h} \times L_{M-R|h}).
\]

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

\[
L_{R|h} = \prod_{r=1}^{R} \prod_{t=1}^{T} \prod_{j=0}^{J} (s_{rjt})^{y_{rjt}} \cdot f(\theta_{1|\bar{R}}, D) \cdot d\theta_{1}.
\]

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,

\[
L_{M-R|h} = \prod_{t=1}^{T} \left[ \sum_{Z_{0t}, \ldots, Z_{Rt}} \int_{0}^{1} \left[ \sum_{r=1}^{R} s_{rjt} f(\theta_{r|\bar{R}}, D) \cdot d\theta_{r} \right] Z_{r}^{y_{rjt}} \right],
\]

where

\[
Z_{r} = O_{j} - \sum_{r=1}^{R} y_{rjt}, \quad \text{and}
\]

\[
\begin{pmatrix}
M-R \\
Z_{0t}, \ldots, Z_{Rt}
\end{pmatrix}
\]

is the multinomial coefficient.

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

\[
\int s_{rjt} f(\theta_{r|\bar{R}}, D) \cdot d\theta_{r} = \sum_{r=1}^{R} s_{rjt}/R.
\]

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 \( L_{Sij} f(\theta_{i|\bar{R}}, D) \cdot d\theta_{i} \) 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

\(^2\)In 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 \( \theta \) 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

\[
 f(y_t, \{\theta_j\}, \delta|S_t, \{x_{n_u}\}) \propto \prod_{t=1}^{T} \left( f(S_t|y_t, \theta_1, \ldots, \theta_K) \right) \prod_{r=1}^{R} \left( f(y_{rt}|\theta_r, x_r) f(\theta_r, \delta, D) f(\theta, D) \right),
\]

where \( y_t \) is the vector of \( y_{rjt} \), f(\( S_t|y_t, \theta_1, \ldots, \theta_K \)) = \( L_{M-Rh,t} \) f(\( y_{rt}|\theta_r, x_r \)) = \( \prod_{j=0}^{J} (s_{rjt})^{y_{rjt}} \) is the choice probability at time \( t \) of cluster \( r \) whose choices are augmented, f(\( \theta_r|\delta, D \)) is the heterogeneity distribution, and f(\( \theta, D \)) 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 \( y_{rjt} \) is proportional to f(\( S_t|y_t, \theta_1, \ldots, \theta_K \))f(\( y_{rt}|\theta_r, x_r \)). Because f(\( y_{rt}|\theta_r, x_r \)) = \( \prod_{j=0}^{J} (s_{rjt})^{y_{rjt}} \), we can generate the candidate draws of \( y_{rt} \) (\( r = 1, \ldots, R \)) 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 \( s_{rjt} (j = 0, \ldots, J) \), which is the logit probability. The candidate draws of \( y_{rt} \) are qualified if the resultant \( Z_{rt} \) 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 \( y_{rt} \) is obtained, the acceptance probability of this candidate draw is given by the ratio of f(\( S_t|y_t, \theta_1, \ldots, \theta_K \)) calculated from the current draw and the one calculated from the \( y_{rt} \) 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, \( \theta_r \) can be generated using a random walk chain Metropolis–Hasting algorithm. Routines for generating the hyperparameters (\( \delta, D \)) 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 as

\[
 u_{ijt} = \varepsilon_{ijt}. 
\]

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 \( \theta \)'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 \( \overline{\theta} \) and heterogeneity matrix \( D \) take the following values:

\[
 \overline{\theta} = (1, 1, -1, .5, \text{and}) 
\]

\[
 D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & .5 \end{bmatrix} 
\]

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

\[
 S_{jr} = \frac{\sum_{t=1}^{5000} y_{ijt}}{5000} \quad (\text{for all } j \text{ and } t). 
\]

The estimation task is to estimate \( \overline{\theta} \) and \( D \) on the basis of the simulated share \( S_{jr} \). If the method is valid, we should get...
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 $\theta$ as $(0, 0, 0, 0, 0)$ and for the four diagonal elements in $D$ as $0.1$. The starting values for $y_{it}$ are randomly generated with an equal choice probability of the three alternatives at the individual level, given the consistency condition ($\Sigma_{y_{it}} \leq y_{jt}$ 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 $R$. 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 $\theta_i$’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 sensitive 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.

### 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.

<table>
<thead>
<tr>
<th>True Values</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$ (Brand 1)</td>
<td>1.0 1.06 (.12)</td>
</tr>
<tr>
<td>$\theta_2$ (Brand 2)</td>
<td>1.0 1.06 (.12)</td>
</tr>
<tr>
<td>$\theta_3$ (Price)</td>
<td>-1.0 -1.03 (.11)</td>
</tr>
<tr>
<td>$\theta_4$ (LAST)</td>
<td>.5 .48 (.11)</td>
</tr>
<tr>
<td>$D_{1.1}$</td>
<td>1.0 1.02 (.30)</td>
</tr>
<tr>
<td>$D_{2.2}$</td>
<td>1.0 1.14 (.37)</td>
</tr>
<tr>
<td>$D_{3.3}$</td>
<td>1.0 1.06 (.19)</td>
</tr>
<tr>
<td>$D_{4.4}$</td>
<td>.5 .64 (.16)</td>
</tr>
</tbody>
</table>
Outside good: 0.72 (.10) N.A. N.A.
Tropicana: 0.10 (.07) 3.75 (.42) .41 (.33)
Dominick’s: 0.11 (.07) 2.39 (.28) .38 (.37)
Minute Maid: 0.07 (.08) 3.35 (.40) .33 (.37)

Table 2

<table>
<thead>
<tr>
<th>Brand</th>
<th>Choice Share</th>
<th>Average Price (cents/ounce)</th>
<th>Promotion Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minute Maid</td>
<td>.07 (.08)</td>
<td>3.35 (.40)</td>
<td>.33 (.37)</td>
</tr>
<tr>
<td>Dominick’s</td>
<td>.11 (.07)</td>
<td>2.39 (.28)</td>
<td>.38 (.37)</td>
</tr>
<tr>
<td>Tropicana</td>
<td>.10 (.07)</td>
<td>3.75 (.42)</td>
<td>.41 (.33)</td>
</tr>
<tr>
<td>Outside good</td>
<td>.72 (.10)</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Notes: N.A. = not applicable.

Table 3

<table>
<thead>
<tr>
<th>Coefficients for</th>
<th>Model 1 (Logit Model Without Dynamics)</th>
<th>Model 2 (Logit Model with Dynamics)</th>
<th>Model 3 (Nested Logit Model with Dynamics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand Choice Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minute Maid</td>
<td>.92</td>
<td>1.15</td>
<td>.83</td>
</tr>
<tr>
<td>Dominick’s</td>
<td>(.16)</td>
<td>(.19)</td>
<td>(.28)</td>
</tr>
<tr>
<td>Tropicana</td>
<td>.31</td>
<td>.44</td>
<td>-.04</td>
</tr>
<tr>
<td>Price</td>
<td>-.19</td>
<td>-.29</td>
<td>-.49</td>
</tr>
<tr>
<td>Promotion</td>
<td>.60</td>
<td>.69</td>
<td>.85</td>
</tr>
<tr>
<td>Last</td>
<td>N.A.</td>
<td>-1.54</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Notes: CV = \ln[\Sigma \exp(V_{ij})]. 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 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 (Guagnoli 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. Our simulation 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.

**Table 4**

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

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (Logit Model Without Dynamics)</th>
<th>Model 2 (Logit Model with Dynamics)</th>
<th>Model 3 (Nested Logit Model with Dynamics)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brand Choice Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minute Maid</td>
<td>.48 (13)</td>
<td>.35 (11)</td>
<td>.41 (14)</td>
</tr>
<tr>
<td>Dominick’s</td>
<td>.69 (22)</td>
<td>.42 (13)</td>
<td>.63 (22)</td>
</tr>
<tr>
<td>Tropicana</td>
<td>.54 (15)</td>
<td>.49 (18)</td>
<td>.60 (22)</td>
</tr>
<tr>
<td>Price</td>
<td>.18 (03)</td>
<td>.14 (03)</td>
<td>.22 (05)</td>
</tr>
<tr>
<td>Promotion</td>
<td>1.71 (42)</td>
<td>.80 (23)</td>
<td>2.70 (77)</td>
</tr>
<tr>
<td>Last</td>
<td>N.A. (283)</td>
<td>13.51 (283)</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

| **Purchase/No-Purchase Model** |                                |                                     |                                           |
| Category value        | N.A.                               | N.A.                                | .09 (01)                                  |
| Last                  | N.A.                               | N.A.                                | 8.40 (203)                                |

Notes: CV = ln[Σ exp(Vij)]/∑ exp(Vij), 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 \((y_t, t = 1, \ldots, T)\).

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

\[
P(y_t | \cdot) \propto \prod_{t=1}^{T} \prod_{j=0}^{J} \left( \prod_{r=1}^{R} s_{jt}^{y_{jt}} \right)^{Z_{jt}} \prod_{r=1}^{R} \prod_{j=0}^{J} y_{jt}^{\gamma_{jt}}
\]

where

\[
s_{jt} = \sum_{k} \exp(\theta_k' x_{jt})
\]

and \(t + 1, \ldots, t'\) 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 \(y_{rt}, r = 1, \ldots, R\), 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 \(s_{jt}(j = 0, \ldots, J)\), 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 \(y_{rt}, r = 1, \ldots, R\), is generated until the resultant \(y_{rt}\) qualifies. Let \(y_{(p)}\) denote the previous draw; then, the next draw \(y_{t(n)}\) is given by

\[
y_{t(n)} = y_{(p)} + \Delta,
\]

with the accepting probability \(\alpha\) given by

\[
\alpha = \min \left\{ \frac{\exp(-1/2(\theta_n - \bar{\theta})'D^{-1}(\theta_n - \bar{\theta}))}{\exp(-1/2(\theta^{(p)} - \bar{\theta})'D^{-1}(\theta^{(p)} - \bar{\theta}))}, 1 \right\},
\]

and \(\Delta\) is a draw from the density \(\text{Normal}(0, .015I)\), where \(I\) is the identity matrix.

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

The posterior distribution of \(D_{k,k}\) is inverted gamma

\[
f(D_{k,k} | \theta_k, \bar{\theta}_k) \propto \text{Inverted gamma}(a, b),
\]

where

\[
a = s_0 + R/2 (s_0 = 3) \quad \text{and} \quad b = \frac{2}{\sum_{r=1}^{N} (\theta_k - \bar{\theta}_k)^2 + 2q_0}
\]

4. Generate \(\bar{\theta}\).

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

where

\[
v = \Psi(\Sigma \bar{\theta}_k + D_0^{-1} \theta_0),
\]

\[
\Psi = (D_0^{-1} + RD^{-1})^{-1},
\]

\(\theta_0 = (0, 0, \ldots, 0)'\), and \(D_0 = 100I\).

REFERENCES


