PREDICTING RETAIL CUSTOMERS' SHARE-OF-WALLET USING SHOPPER LOYALTY CARD DATA[†]

Edward J. Fox (Southern Methodist University)*

Jacquelyn S. Thomas (Northwestern University)**

April 2006

* Assistant Professor, Edwin L. Cox School of Business, Southern Methodist University, Dallas, TX; phone: 214-768 3943; TUefox@mail.cox.smu.edu UT

** Associate Professor, Medill School of Journalism, Northwestern University, Evanston, IL; phone: 214-768 3943; jakki@northwestern.edu

[†] The authors would like to thank John Semple for his comments and suggestions. Any mistakes or omissions are the sole responsibility of the authors.

PREDICTING RETAIL CUSTOMERS' SHARE-OF-WALLET USING SHOPPER LOYALTY CARD DATA

Abstract

Most leading grocery retailers use shopper loyalty cards (also called frequent shopper cards) to gather information about their customers with the objective of determining who should receive targeted marketing offers. However, retailers have not been able to use loyalty card programs to assess customer loyalty. In this paper, we propose and test a hierarchical Bayesian approach to the prediction of customer share-of-wallet loyalty that: (1) models household spending at competing retailers using multi-outlet panel data, (2) applies parameter estimates from the model to the retailer's loyalty card data to develop expectations of customer spending at competing retailers, then (3) predicts share-of-wallet using these spending expectations and observed spending at its own stores. Our approach offers a high degree of predictive accuracy and discriminates well between loyal and non-loyal customers. Among the types of information that retailers may gather with loyalty card programs, we find that geographic information (i.e., travel time to the store and retail concentration around that store) makes a much greater predictive contribution than customer demographics. Finally, we show how retailers can use the posterior marginal distribution of customer share-of-wallet to more profitably determine which customers should receive marketing offers that target loyal customers.

(keywords: retail, shopping behavior, customer relationship management, share-of-wallet, loyalty)

INTRODUCTION

Perhaps the most important marketing innovation among grocery retailers during the past decade has been the shopper loyalty card program. Eighty-nine percent of leading grocery retailers offer shopper loyalty cards, also known as frequent shopper cards (Progressive Grocer 2001). Retailers issue the cards to their customers, who must present them to qualify for discounts, coupons, free goods, etc. Consumers believe that shopper loyalty cards offer important savings opportunities, and use them more than any other offers from grocery retailers, including private label products (Food Marketing Institute 2004).

According to *Food Marketing Institute's Guide to Planning Frequent Shopper Programs* (Willard Bishop Consulting 1995), retailers benefit from card programs because they are able to improve their understanding of shopper behavior (p. 7), monitor shopper purchases (p. 14), "segment shoppers to identify and reward best customers" (p.14), and target shoppers effectively with mailings, coupons, and other promotions (p. 15). Why the focus on understanding and managing individual shoppers? Because "…building upon the 'share of customer' is now viewed as more meaningful and possibly more controllable than simply focusing on market share as a strategic objective" (p. 7). Shopper loyalty card programs have become the most widely-used tool for building customer share-of-wallet [SOW] (as share-of-customer is more commonly known), the prevailing metric of shopper loyalty.

However, retailers have not been able to use loyalty card programs to *assess* customer loyalty. This is because SOW depends on shoppers' spending at competing retailers, which is not captured in card programs. Yet the capability to assess shopper loyalty would be quite valuable, because loyalty is a useful predictor of how customers will respond to marketing offers. Moreover, customer loyalty is increasingly viewed as an important indicator of retailer performance. For example, hotels (Noone, Kimes, and Renaghan 2003) and banks (Jarrar and Neely 2002; Garland 2003; Mittal 2004) have begun to use customer SOW to measure their effectiveness, particularly in cross-selling efforts, while firms in the apparel industry are now using SOW in such marketing decisions as the development of merchandising tactics and the opening of new stores (Huff 2002).

The objective of this paper is to introduce a new statistical approach that uses shopper loyalty card data to predict SOW, enabling retailers to select customers for targeted marketing offers more profitably. The approach involves: (i) first estimating models of customer spending at competing retailers using syndicated multi-outlet panel data, (ii) applying the model parameters to shopper card data in order to generate SOW predictions, then (iii) classifying customer loyalty, based on SOW, to determine who should receive targeted marketing offers. In essence, SOW predictions are generated by modeling panel data, then applying parameter estimates from the model to shopper loyalty card data as one would a holdout sample. The panel data are modeled using a multivariate system of hierarchical Bayesian Tobit censored regressions, with customer spending at each retailer (i.e., store chain) as the dependent variable. The model is estimated using the Gibbs Sampler.

Several factors favor our approach. The Bayesian hierarchy captures systematic (as well as random) sources of heterogeneity in consumer spending, permitting us to determine the predictive contributions of different types of information captured in card programs: (i) purchase histories, (ii) demographic, and (iii) geographic variables. Also, because the Gibbs Sampler can be used to sample from the posterior distribution of any function of model parameters, we are able to construct Bayesian prediction intervals for each customer's SOW. Assuming that response to a targeted marketing offer depends on the loyalty of customers who receive it, the prediction intervals can be used to select customers to receive that offer. Should the retailer select customers if there is a 10% probability that they are loyal? 50%? 90%? We demonstrate how retailer profits can be increased by selecting customers based on the posterior probability that they are loyal. In addition, because consumer spending at competing retailers is modeled explicitly, we could determine how much of a customer's wallet is going to non-traditional competitors such as Wal-Mart supercenters and discount stores, and why. Finally, our approach uses only shopper loyalty card data, which most grocery retailers already gather, and syndicated multi-outlet panel data, which is widely available and easily procured.

The next section of this paper reviews the relevant literature. In section 3, we present a brief overview of the data used to estimate our model. Section 4 presents the model and discusses how it is estimated. The empirical results from that estimation are discussed in section 5. To demonstrate the value of our research for managers, we show how to profitably select customers for targeted marketing offers in section 6. The final section of the paper discusses the limitations of our research and suggests avenues for future research.

LITERATURE REVIEW

As SOW has come to be viewed as an important indicator of firm performance, a stream of research has developed that focuses on the antecedents of SOW. Customer satisfaction (Bowman and Narayandas 2001; Magi 2003; Keiningham, Perkins-Munn, and Evans 2003; Verhoef 2003), affective commitment (Verhoef 2003), shopper characteristics (Macintosch and Lockshin 1997; Magi 2003), product or service quality (Odekerken-Schroder et al. 2001), loyalty programs (Magi 2003; Verhoef 2003), and direct mail (Verhoef 2003) have all been studied as antecedents of SOW (Magi 2003; Keiningham, Perkins-Munn, and Evans 2003). However, the weak relationships found between SOW and the proposed antecedents have led researchers to question the measurement of SOW (Dowling 2002; Magi 2003; and Verhoef 2003). In general, SOW has been measured by either consumer self reports (e.g., Bowman and Narayandas 2001; Odekerken-Schroder et.al. 2003; Garland 2003; Magi 2003), or consumer self reports in combination with a firm's customer data (e.g., Verhoef 2003). Interestingly, Keiningham, Perkins-Munn, Evans (2003) compared self-reported measures of SOW with measures of repurchase intentions and concluded that self-reported measures may not be reliable.

Perhaps in response to these concerns, research has emerged that focuses on the prediction of customer SOW. Unfortunately, publications in this area are sparse and do not always focus on SOW, per se. For example, Rust, Lemon and Zeithaml (2004) present an approach that leverages Markov switching matrices to predict customer lifetime value, a discounted estimate of the customer's future dollar contributions to a firm. From customer lifetime value, the authors derive what they refer to as customer equity share, a similarly forward-looking measure of relative customer value across brands or firms. Though it is not SOW, customer equity share is similar in spirit.

Chen and Steckel (2005) use Markov switching matrices to predict SOW. Using credit card data from a single provider, they estimate customer interpurchase times, which are used to infer the number of purchases that customers had made from competing providers. Unlike other research in the area, they define SOW based on the number of purchases made by a customer, not his/her expenditures.

The research most closely related to ours is that of Du, Kamakura, and Mela (2005). They propose a multivariate factor analytic model to predict SOW[1]. As in our approach, Du, Kamakura, and Mela augment the firm's internal customer data with external data. The internal data came from a bank, a sample of whose customers were also surveyed about their business in ten different product categories at competing financial service providers. Total outside balances and holdings in each category fulfill the external data requirement. Du, Kamakura, and Mela model customer share-of-requirements in each of the ten categories, which are used to predict total SOW. In contrast, we predict SOW by modeling consumer spending separately at each competing retailer, including Wal-Mart discount stores, supercenters, and Sam's Club stores. Modeling spending at competitors, especially Wal-Mart, is particularly relevant for grocery retailers because shoppers increasingly buy their groceries at supercenters and mass merchandisers (29% in 2003 vs. 26% in 2000), as well as warehouse clubs (17% in 2003 vs. 14% in 2000, FMI 2004 p. 22). At the same time, the proportion of shoppers who buy categories such as cereal, over-the-counter drugs, and paper products at supermarkets has decreased by more than ten percent (FMI 2004 p. 24). By modeling competing retailers individually, our approach could also be used to address specific questions about the character of retail competition and *how* the customer's wallet is being allocated.

[1] DKM estimate both SOW and total wallet, or total amount spent in categories offered by the firm. Note that our method also estimates total wallet as the denominator of SOW, though it is not examined in depth in this paper.

DATA

The dataset used in this study is from a multi-outlet panel in a major metropolitan market in the southeastern United States during the period September 2002 through September 2004. Panelists recorded all of the purchases they made at retailers selling packaged goods products that could be identified with a Uniform Product Code (UPC). Unlike panel datasets typically used for marketing research, purchases were recorded not only at grocery stores but at other retailers, including mass merchandisers, supercenters, warehouse clubs, drug and dollar stores. Our analysis focuses on the top seven retailers in the market which together account for over 85% of all consumer spending on products identified by UPC at known retailers. Of particular interest, three Wal-Mart formats are among the top seven retailers.

Table 1 presents descriptive information about these seven retailers, including average quarterly consumer spending and number of store visits, percent of households shopping there (penetration), and average distance from each panelist to the closest store of that chain, a measure of spatial convenience. As the table shows, the retailers include four grocery store chains, Wal-Mart supercenter and "division one" discount stores[2], and Sam's Club, Wal-Mart's warehouse club format. A higher percentage of households visit Grocery 2 in each quarter than any other chain (79%), followed by Wal-Mart Supercenter (62%) and Grocery 3 (57%). Sam's Club (30%) and Wal-Mart Discount (34%) have the lowest household penetration. The average number of store visits follows a similar pattern, with the more trips made to Grocery 2 (7.1/quarter) than to any other chain, followed by Wal-Mart Supercenter (4.0/quarter) and Grocery 3 (3.7/quarter). Consumers also spend the most money on packaged goods at Grocery 2 ($184/quarter), followed in order by Grocery 3 ($145/quarter) and Wal-Mart Supercenter ($122/quarter). Consumers spend less at Wal-Mart Discount ($30/quarter) and Sam's Club ($33/quarter)

[2] though both formats carry the Wal-Mart brand name, we separate them because they offer a substantially different mix of products

than at other chains. Taken together, these aggregate shopping data suggest that failing to include Wal-Mart's formats, in particular its supercenters, could bias our analysis of SOW.

< Insert Table 1 about here >

We were concerned that some panelists may not have faithfully recorded purchases in all outlets over the entire 104-week period. To avoid possible biases due to haphazard recording, we eliminated panel households that did not meet both of the following criteria: (i) purchases were recorded in each of the 24 consecutive months during which data were gathered, and (ii) purchases recorded by the household averaged at least $25 per week. Though these screening criteria may have resulting in discarding a few households that faithfully recorded their purchases (e.g., single-person households with very low disposable incomes or households that took extended vacations), we believe this is less problematic than including households that systematically under-reported their spending. We also eliminated households for which we did not have precise locations (nine-digit zip codes) as well as the few households that made most of their purchases outside of the seven retailers under study.

The resulting dataset includes 359 households that made an average of 71 shopping trips and spent an average of $3215 on packaged goods at the top seven retailers during the 24-month duration of the data. Table 2 shows descriptive statistics for these households.

< Insert Table 2 about here >

We divided the 359 households randomly into an estimation sample of 211 households and a 148-household holdout sample. Holding out entire households rather than individual spending observations is consistent with our objective of applying the model estimates to members of a retailer's shopper loyalty card program. In this application, we use the holdout sample to assess the predictive accuracy that we could expect from each retailer's shopper loyalty card data. Statistical testing of the means of demographic and geographic variables between the two samples showed no significant differences. Finally, the first nine months of data were used to measure shopping behavior variables (predictors in the model which will be described in the next section) for each household. We use the remaining fifteen months of data to construct the estimation and holdout samples.

MODEL

Recall that our objective is to predict SOW for packaged goods spending outside of our estimation sample; i.e., among the members of a focal retailer's shopper card program. To do so, we model packaged goods spending at competing retailers' stores (and the focal retailer's own stores) using only information available from its shopper card program. Specifically, for every household, we jointly model the probability of shopping at each retailer during a given time period and spending during that period, conditional on shopping. Note that our model specifies that spending be aggregated over time (and store visits) to improve out-of-sample prediction of SOW, which is a temporally aggregate measure (Man 2004).[3] Quarterly spending observations are specified because this level of temporal aggregation is long enough that most observations are non-zero, yet short enough that we have repeated spending measures to reliably estimate individual-level parameters. Across estimation and holdout samples, 49% of quarterly spending observations at the seven retailers are zeros, indicating that an average household shops at just over half of the available store chains. Given the large number of zero observations, Tobit censored regression models are appropriate for this application (Tobin 1958; see Fox, Montgomery, and Lodish 2004 for application of Tobit models to shopping behavior).

We have chosen not to model SOW explicitly, because we do not observe SOW in shopper loyalty card data. Instead, we predict SOW by modeling consumer spending at all retailers, conditioning spending predictions at other retailers on observed spending from the focal retailer's shopper loyalty card data, and compute SOW from these quantities.

---

[3] Because the underlying data generating process of consumer spending on store visits is known to complex and highly nonlinear, disaggregate spending models would almost certainly suffer from misspecification.

The dependent variable of the Tobit model is y^*_{hrq} where h=1,…, H indexes households, r=1,…, R indexes retailers (store chains), and q=1,…, Q indexes quarterly observations. The dependent variable, y^*_{hrq}, takes on the value of household h's spending at retailer r during quarter q if any spending is observed; otherwise it is a latent value less than or equal to zero. The observation equation is therefore

$$y_{hrq} = \begin{cases} y^*_{hrq} & \text{if } y^*_{hrq} > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

The latent dependent variable is modeled with a linear regression

$$y^*_{hrq} = \alpha_{hr} + \boldsymbol{\theta}'_r \mathbf{x}_q + \varepsilon_{hrq} \tag{2}$$

where α_{hr} is a household-specific intercept, $\mathbf{x}_q$ is a vector of time-varying factors—quarterly dummies to capture seasonality and a linear trend variable, and $\boldsymbol{\theta}_r$ is the associated vector of coefficients.[4] As the subscripts show, both the intercept α_{hr} and parameter vector $\boldsymbol{\theta}_r$ are specific to the retailer. Thus, we permit quarterly spending at each retailer to be affected differently by household-specific and time-varying factors.

The random error term, ε_{hrq}, is distributed:

$$\varepsilon_{hrq} \sim N(0, \Sigma) \tag{3}$$

where $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1R} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{R1} & \sigma_{R2} & \cdots & \sigma_{RR}^2 \end{bmatrix}$

Following Fox, Montgomery and Lodish (2004), we observe that unmodeled spending at one retailer is likely to be predictive of spending at others. For example, household spending below the predicted amount at Kroger might be associated with spending at Albertsons and/or Safeway above what the model

[4] We also estimated equation (2) with the dependent variable, y^*_{hrq}, log transformed. This alternative specification offered substantially poorer fits and out-of-sample prediction.

predicts. We therefore allow spending residuals to covary across stores, as reflected in the error covariance matrix Σ. To estimate this covariance matrix, we structure the design matrix as in seemingly unrelated regression (Zellner 1962).

Bayesian Hierarchy

The household-specific intercept for retailer r, α_{hr}, is modeled with the hierarchical equation

$$\alpha_{hr} = \mu_r + \boldsymbol{\beta}'_r \mathbf{b}_h + \boldsymbol{\delta}'_r \mathbf{d}_h + \boldsymbol{\gamma}'_r \mathbf{g}_{hr} + \zeta_{hr} . \tag{4}$$

This equation includes both systematic and random sources of heterogeneity in household spending. Systematic sources of heterogeneity are captured with an intercept term and three vectors of predictors: (i) variables that reflect shopping behaviors at the focal retailer, $\mathbf{b}_h$, (ii) household demographics, $\mathbf{d}_h$, and (iii) geographic variables that relate household h to the nearest store of retailer r, $\mathbf{g}_{hr}$. The corresponding parameters are μ_r, $\boldsymbol{\beta}_r$, $\boldsymbol{\delta}_r$, and $\boldsymbol{\gamma}_r$, respectively. The three vectors of predictors reflect different types of information that may be available from a shopper loyalty card program. Our analysis will focus on these systematic sources of heterogeneity, because accurate SOW prediction depends on effectively incorporating individual differences. Unmodeled heterogeneity in α_{hr} is reflected in the residual term, ζ_{hr}, which is distributed

$$\zeta_{hr} \sim N(0, \Omega) \tag{5}$$

where $\Omega = \begin{bmatrix} \omega_1^2 & & & \\ & \omega_2^2 & & \\ & & \ddots & \\ & & & \omega_R^2 \end{bmatrix}$.

We allow the variances of this residual term to differ across retailers, reflecting differences in baseline shopping frequencies and basket sizes between retailers. Off-diagonal terms are restricted to be zero.

Shopping Behaviors – The first vector of predictors, $\mathbf{b}_h$, is comprised of shopping behaviors that can be measured using customer purchase histories. Note that there is no retailer subscript in $\mathbf{b}_h$, because we measure behaviors using the purchase histories only at the focal retailer, which we denote r'. This is the retailer from which shopper loyalty card data is available. When model estimates are subsequently applied to members of the card program at retailer *r′*, purchase histories there will be the only source of behavioral data.

We allow the behavior parameters, β_r, to be retailer-specific, because behaviors at one retailer's stores can be predictive of different spending patterns at competing stores. Consider the four specified behavioral measures: (i) *the percentage of household purchases made on discounted items* and (ii) *feature advertised items* are indicative of the propensity to search for low prices in-store or prior to shopping, and perhaps also of a preference for stores that offer low prices; (iii) *average spending per trip* is indicative of a household's opportunity cost (small baskets suggest low opportunity costs) which could differentially affect store preferences; and (iv) *average package size* of items purchased can reflect both a household's holding costs and price sensitivity (larger package sizes are generally offered at lower unit prices), both of which could affect store preference.[5] The selection of these four shopping behavior variables is admittedly ad hoc.[6] We include them in part for demonstration—a more systematic approach to selecting shopping behavior variables is left for future research. Recall that the first nine months of available data, which were used to measure shopping behavior variables, are not included in the estimation and holdout samples. Thus, endogeneity is avoided.

[5] Because measures of purchase volume differ among categories, we operationalize the average package size of household *h* as an average of category level indices, weighted by the proportion of the household's spending in each category as follows:

$$\overline{q}_h = \frac{1}{y_h} \sum_{c=1}^{C} \frac{y_{hc}\overline{q}_{hc}}{\overline{q}_c}$$

where c=1, …, C indexes categories, $\overline{q}_{hc}$ is the average package size of purchases by household h in category c, and y_{hc} is the total of household h's expenditures in category c.

[6] Our variable selection was made in consultation Gerry Usrey, former Vice President of Retail Insights R&D at PEPSICO.

Demographics – The second vector of predictors in the hierarchical equation, $\mathbf{d}_h$, captures household demographics: (i) *family size* (# members), (ii) *household income* (x$10,000), (iii) *head-of-household age* (x10), (iv) head-*of-household education* (college degree=1, no college degree=0), and (v) *working woman* in the household (yes=1, no=0).[7] We expect the effect of demographics on spending to depend on the retailer's positioning and marketing mix, so the parameters are allowed to be retailer-specific. For example, everyday-low-price retailers (Grocery 2 and the three Wal-Mart formats in this application) tend to offer lower average prices, so they appeal to households with lower incomes and larger families (Bell, Ho, and Tang 1998; Fox, Montgomery, and Lodish 2004). Households with higher incomes and smaller families tend to prefer more convenient, but higher priced, retailers.

Household demographics can be gathered in either of two ways. Either the retailer can ask consumers who apply for a shopper loyalty card to disclose their income, family size, age, etc., or demographics can be appended from syndicated databases which compile data from the census and other sources.[8]

Geographic Variables – The last vector of predictors in the hierarchical equation, $\mathbf{g}_{hr}$, captures (i) *travel time* from household *h* to the closest store of retailer *r*, and (ii) *retail concentration* around that store, which we define as the number of other stores to which one could travel in five minutes or less.[9] These variables capture the propensity of household *h* to spend at retailer *r* due to location.[10] Because response to travel time (Fox, Montgomery and Lodish 2004) and retail concentration (Miller, et al. 1999) depend

[7] We also estimated model specifications which included home ownership and head-of-household age among the demographic variables, but they offered little additional explanation of household spending.

[8] Providers of syndicated demographic data include such companies as ESRI, Axciom and MapInfo.

[9] Retail concentration is intended to measure the number of retail outlets nearby which sell packaged goods products. We have information about 895 such stores which together comprise all grocery stores, drug stores, mass merchandisers, supercenters, warehouse club stores and dollar stores in the market.

[10] The locations of all stores in any market are available from geographic data provides such as ESRI, MapInfo and Claritas.

on the retailers' format and positioning, we allow these parameters to vary by retailer. For both geographic variables, we posit a diminishing marginal effect. For example, the marginal effect on consumer spending of ten additional minutes of travel will be greater if the change is from five to fifteen minutes than if the change is from 45 to 55 minutes. We confirmed this intuition by experimenting with concave transformations (log(·), sqrt(·), and -1/(·)) as well as untransformed geographic variables. The square root transformation resulted in the best fits, and was therefore applied to the travel time (i.e., sqrt(minutes)) and retail concentration (i.e., sqrt(1 + # of stores)) variables. We expect travel time to be negatively related to household spending at the retailer. Retail concentration around a store can affect household spending there either positively or negatively, depending upon whether the positive effect of symbiosis with nearby stores is greater or less than the negative effect of competition with those stores (Miller, et al. 1999).

In summary, systematic differences between individuals' propensity to spend at a given retailer depend on three sets of variables. These variables reflect observed shopping behaviors, demographics, and the relative locations of consumers and stores. In the empirical analysis that follows, we will examine the predictive contribution of each set of variables.

Specification of Priors

To ensure that posterior distributions are defined, we specify prior distributions for parameters and variances/covariances. The first stage of our model is given by equations (2) and (3). To complete this stage, we specify a prior for the parameters

$$\theta \sim N\left(\bar{\theta}, \nu_\theta V_\theta\right) \tag{6}$$

where $\theta = \left[\theta_1' \quad \theta_2' \quad \cdots \quad \theta_R'\right]'$, and a natural conjugate prior for the error covariance matrix

$$\Sigma^{-1} \sim \text{Wish}\left(\nu_\Sigma, V_\Sigma^{-1}\right). \tag{7}$$

To complete the second stage of our model, given by equations (4) and (5), we specify a prior distribution for the elements of the diagonal covariance matrix, Ω,

$$\omega_r^{-2} \sim \chi^2\left(\nu_\omega, \overline{\omega}_r^{-2}\right). \tag{8}$$

To complete the third stage of our model, we specify a prior distribution for the hyper-parameters in the hierarchical equation (4). First, we rewrite the hierarchical equation as

$$\alpha_{hr} = \boldsymbol{\psi}_r' \mathbf{z}_{hr} + \zeta_{hr} \tag{9}$$

where $\mathbf{z}_{hr} = \begin{bmatrix} 1 & \mathbf{g}_{hr}' & \mathbf{d}_h' & \mathbf{b}_h' \end{bmatrix}'$ and $\boldsymbol{\psi}_r = \begin{bmatrix} \mu_r & \boldsymbol{\gamma}_r' & \boldsymbol{\delta}_r' & \boldsymbol{\beta}_r' \end{bmatrix}'$.

We assume a normally distributed prior for $\boldsymbol{\psi}$ with

$$\boldsymbol{\psi} \sim N\left(\overline{\boldsymbol{\psi}}, \nu_\psi V_\psi\right) \tag{10}$$

where $\boldsymbol{\psi} = \begin{bmatrix} \boldsymbol{\psi}_1' & \boldsymbol{\psi}_2' & \cdots & \boldsymbol{\psi}_R' \end{bmatrix}'$.

We select relatively diffuse priors in equation (10) so that posteriors will be determined almost entirely by the data: $\overline{\boldsymbol{\psi}}$ = [**0**], $\nu_\psi = 1$, and $V_\psi = 10^5 I$. We determine the values of $\overline{\boldsymbol{\theta}}$, V_θ, V_Σ and $\overline{\omega}_r^2$ empirically. To set these priors, we first estimated independent Tobit models for each retailer, pooling across households. These pooled Tobit models, which will subsequently be used as the baseline model in goodness-of-fit testing, are jointly specified by equation (1),

$$y_{hrq}^* = \alpha_r + \boldsymbol{\theta}_r' \mathbf{x}_q + \varepsilon_{hrq}, \text{ and} \tag{11}$$

$$\varepsilon_{hrq} \sim N\left(0, \sigma_r^2\right). \tag{12}$$

We set $\overline{\boldsymbol{\theta}}$ to the parameter estimates from these pooled Tobit models. Following Montgomery (1996), we incorporate scaling constants in setting priors for the variances: V_θ, V_Σ, and $\overline{\omega}_r^2$. These scaling constants reflect the expected reduction in variance from the pooled Tobit models. For example, because we model several sources of systematic heterogeneity across customers, our prior on the residual variance term in the hierarchical equation, $\overline{\omega}_r^2$, should be much less than the variance of α_r from the pooled Tobit model for retailer r. We therefore specify $\overline{\omega}_r^2$ using the variance of α_r from the pooled Tobit model, $s_{\omega,r}^2$, and a scaling constant, k_ω,

$$\overline{\omega}_r^2 = k_\omega^2 s_{\omega,r}^2 \tag{13}$$

Setting $k_\omega = .75$ would correspond to a prior belief that the residual standard deviation of the hierarchical equation, ω_r, is 75% of standard deviation of α_r from the pooled Tobit model. Using the same approach for V_θ and V_Σ, we set those prior distributions as follows:

$$V_\theta = \text{diag}\left[k_\theta^2 s_{\theta,1}^2 \quad k_\theta^2 s_{\theta,2}^2 \quad \cdots \quad k_\theta^2 s_{\theta,R}^2\right] \tag{14}$$

$$V_\Sigma = \text{diag}\left[k_\Sigma^2 s_{\Sigma,1}^2 \quad k_\Sigma^2 s_{\Sigma,2}^2 \quad \cdots \quad k_\Sigma^2 s_{\Sigma,R}^2\right] \tag{15}$$

where $s_{\theta,r}^2$ and $s_{\Sigma,r}^2$ are empirical estimates of the variance of θ_r and the error variance of the pooled Tobit model for retailer r, respectively. We do not have a clear intuition about how much to reduce the empirical estimates for our priors, although we expect predictors in the hierarchical equation to explain most of the variation in individual-level intercepts. Therefore, we set the scaling constants $k_\theta = k_\Sigma = 0.5$ and $k_\omega = 0.25$.

The last values to be chosen are ν_θ, ν_ω, and ν_Σ, which determine how much weight to assign to our empirical priors. Given that individual differences play an important role in SOW prediction and our limited intuition about the scaling constants, it is important not to shrink individual parameter estimates toward the priors too much. We therefore set $\nu_\theta = \nu_\omega = \nu_\Sigma = 5$. To determine how much the specified priors affect the posterior distributions, we tried doubling ν_θ, ν_ω, and ν_Σ. Neither inferences nor predictions based on the posteriors were significantly affected.

Estimation

We estimate the latent dependent variable, y_{hrq}^*, using data augmentation (Tanner and Wong 1987) which has been adapted for use in Tobit models (Wei and Tanner 1990; Chib 1993). If the dependent variable is not observed (i.e., $y_{hrq} = 0$), a value of the latent dependent variable is imputed from a truncated normal distribution so that $y_{hrq}^* \leq 0$ (See Appendix for details).

Equations (1)-(5) are estimated using the Gibbs Sampler. The Gibbs Sampler involves sampling sequentially from all relevant conditional distributions over a large number of iterations. Specifications of the full conditional distributions of model parameters and variances, along with the latent dependent variable, are detailed in Appendix. We make 50,000 draws from a single continuous Gibbs chain. The first 25,000 draws are used as a "burn in" period and subsequently discarded. We confirm convergence of the Gibbs Sampler by applying Geweke's test (1992). To reduce autocorrelation in the Gibbs draws, we "thin the line," using every 10^{th} draw for inference.

EMPIRICAL RESULTS

Because the model uses information from a grocery retailer's shopper loyalty card program, we estimate separate models for each of the four grocery retailers in our dataset. Each model uses shopping behaviors at one of the grocery retailers as predictors, and spending at that retailer to condition spending expectations at other retailers. The four models are otherwise identical.

Goodness-of-Fit

***Consumer Spending Model* –** We begin the empirical analysis by comparing the fit of the specified model to a baseline model and other selected nested models. The selected nested models each omit one of the three types of information that can be captured in shopper loyalty card programs—shopping behaviors, demographics, or geographic variables. By comparing the nested models to the full specification, we are able to quantify the incremental contribution to fit of each type of information.

Table 3 shows in and out-of-sample fit statistics for (i) four full model specifications (each using information from one of the grocery retailers in our dataset), (ii) four comparable specifications without geographic variables, (iii) four specifications without demographics, (iv) a single specification without shopping behavior variables, and (v) the baseline model. Each nested model may be compared to the full model that uses information from the same retailer. The model without shopping behaviors and the baseline model are comparable to all four full model specifications.

We begin with the baseline model: independent Tobits for each retailer, pooled over households.

Recall that this pooled model, which was used to set empirical priors, is jointly specified by equations (1), (11), and (12). In the bottom row of Table 3, we see that the baseline model has a log-likelihood of –28,528 in sample and –19,706 out-of-sample.[11] Log-likelihoods of the four full models in the top panel of the table are much higher, ranging from a low of –18,110 in sample and –15,108 out-of-sample (using information from Grocery 1) to a high of –17,723 in sample and –14,458 out-of-sample (using information from Grocery 3). Clearly, our multivariate system of hierarchical Bayesian Tobit models fits the data much better than the baseline model.

< Insert Table 3 about here >

Now we consider the other nested models. In-sample fits are compared using two criteria well suited for Bayesian models. The first is log-marginal density, computed using the modified LaPlace method (Kass and Raftery 1995). The second criterion is the Deviance Information Criterion, or DIC (Speigelhalter et al. 2002). This relatively new measure of fit was proposed for Bayesian models estimated with MCMC methods as an analog of the information criteria commonly applied to classical statistical models. Out-of-sample fits are assessed by comparing log-likelihoods of the holdout sample, computed by setting model parameters to their posterior means.

In sample, we find that the full specification is preferred on the basis of both log-marginal density and DIC, regardless of which retailer's information is used. Out-of-sample, among models using information from Grocery 1, 2, and 4, the specification without demographics has the highest log-likelihood. Only among models using information from Grocery 3 does the full specification have the highest log-likelihood (only slightly higher than the specification without demographics). Thus, evidence from the holdout sample suggests that including demographics in the model may "overfit" the data. The nested specification without geographic variables offers the poorest fits (in and out-of-sample) regardless

[11] Goodness-of-fit measures that are more appropriate for Bayesian models are not reported in Table 3 for the baseline model.

of which retailer's information is used. We conclude that geographic variables offer substantial explanation of spending behavior, even in the presence of the other predictors.

Share-of-Wallet Prediction – We now turn to the objective of our research—predicting SOW for a given retailer's customers. The expectation of the latent dependent variable, $E(y^*_{hrq})$, must first be converted to expected customer spending, $E(y_{hrq})$. This computation is well known for univariate Tobit models; we adapt it for our multivariate system of Tobits as follows

$$E\left(y_{hrq}\right) = \Phi\left(\frac{\alpha_{hr} + \boldsymbol{\theta}'_r \mathbf{x}_q}{c_r}\right)\left(\alpha_{hr} + \boldsymbol{\theta}'_r \mathbf{x}_q + \lambda_{hrq} c_r\right) \tag{16}$$

where $\Sigma = CC'$, c_r is the r^{th} diagonal element of the Cholesky root C, and

$$\lambda_{hrq} = \varphi\left(\frac{\alpha_{hr} + \boldsymbol{\theta}'_r \mathbf{x}_q}{c_r}\right) \Big/ \Phi\left(\frac{\alpha_{hr} + \boldsymbol{\theta}'_r \mathbf{x}_q}{c_r}\right)$$

($\Phi(\cdot)$ and $\varphi(\cdot)$ denote the CDF and PDF of the standard normal distribution, respectively). We use diagonal elements of the Cholesky root rather than the square root of the diagonal elements of Σ to exploit error covariances across retailers.

Because we have access to purchase histories from the shopper loyalty card program at retailer r', we observe actual quarterly customer spending at its stores. These observations are used to condition spending expectations at other retailers ($r \neq r'$) as follows

$$E\left(y_{hrq}\right) | \, y_{hr'q} = E\left(y_{hrq}\right) + \left(y_{hr'q} - E\left(y_{hr'q}\right)\right)\left(\sigma_{rr'} \big/ \sigma^2_{r'}\right) \tag{17}$$

The conditional spending expectations can then be used to predict SOW over the duration of the data (q=1, …, Q) as follows

$$SOW_{hr'} = \frac{y_{hr'}}{y_{hr'} + \sum_{r \neq r'} E\left(y_{hr}\right) | \, y_{hr'}} \tag{18}$$

where $y_{hr'} = \sum_{q=1}^{Q} y_{hr'q}$ and $E(y_{hr}) \mid y_{hr'} = \sum_{q=1}^{Q} E(y_{hrq}) \mid y_{hr'q}$.

Goodness-of-fit for SOW predictions of the full and nested models are compared in Table 4 on the basis of mean absolute deviation (MAD). Observe that all models estimated using information from a given grocery retailer have the same number of customers, though that number varies by retailer. This is because each retailer has a different number of customers in the holdout sample of 148 panel households. We assume that only those who visited the retailer sometime during the period of the data would have enrolled in its shopper loyalty card program; others would not be counted among its customers.

< Insert Table 4 about here >

We find that the full model is preferred to all nested models for those using information from Grocery 1, 2, and 4. The nested model without demographics is preferred for Grocery 3, suggesting that demographics may have little predictive power beyond the shopping behaviors of customers at Grocery 3. Observe that nested models without purchase histories predict SOW poorly, because actual spending at retailer *r′* is not observed. We also find that the baseline model is less predictive than the full models using information from each of the four retailers, an average of 6% worse.

To evaluate the overall adequacy of the model, we also compute a pseudo- R^2 measure that reflects the proportion of variance in SOW that is explained. Across the full models using information from each of the four grocery retailers, the average pseudo- R^2 is 0.749 with a range of 0.620 to 0.805.

APPLICATION TO CUSTOMER SELECTION

The effectiveness of our approach to predicting SOW is best evaluated in the context of the purpose for which it was intended—selecting customers for targeted marketing offers. Assuming that expected profits from targeted marketing offers depend on the loyalty of customers receiving them, SOW predictions can be used to select customers to receive those offers. Though any number of selection or classification criteria can be applied, we will apply a simple binary loyalty classification rule: customers who make most of their expenditures (i.e., >50%) at one retailer's stores will be considered loyal to that retailer.[12] The predictive power of our approach is evident when classifying customers according to this simple rule.

< Insert Figure 1 about here >

Figure 1 reports confusion matrices for customer classification using full model specifications with information from each of the four retailers. Recall that each retailer has a different number of customers in our holdout sample. Summing the confusion matrices from left to right yields the number of those customers that are actually loyal and non-loyal. Summing confusion matrices from top to bottom yields the number of customers that are predicted to be loyal and non-loyal. Thus, diagonal elements of the confusion matrices are correct predictions; off-diagonal elements are predicted incorrectly. Across models using information from the four grocery retailers, we correctly identify loyal customers 77.5% of the time. Noting that the base rate of customer loyalty is only 18.8% across the four retailers, we find that loyal customers are predicted at 4.1 times the rate we would expect by chance. By comparison, the pooled Tobit baseline model correctly identifies loyal customers only 32.5% of the time, just 1.7 times the rate we would expect by chance. This comparison illustrates the importance of incorporating heterogeneity in the identification loyal customers. Combining loyal and non-loyal customers, we find that the full model correctly classifies customers 89.2% of the time. Thus, in this application we are able to effectively discriminate between loyal and non-loyal customers.

The customer loyalty classifications reported in Figure 1 were generated using the means of posterior marginal distributions of parameters in the spending models, estimated with the Gibbs Sampler. Allenby and Rossi (1999) demonstrate that pricing decisions made using means of posterior marginal distributions of price response parameters are not necessarily most profitable for the firm. They noted that "…we must use the entire posterior distribution of the household parameters, rather than just point estimates to solve the marketing problem" (p. 75). Because Bayesian models estimated using MCMC

[12] Alternative classification criteria for loyal vs. non-loyal customers might be 60%, 66.7% or 75% SOW. An alternative classification scheme might be loyal vs. switcher vs. competitor loyal.

methods can generate posterior distributions of any function of model parameters, we do so for SOW. Recall that SOW depends on the joint distribution of parameters related to spending at all retailers. Following Allenby and Rossi, we will demonstrate how retailers can use the entire posterior distribution of customer SOW to make more profitable decisions than they would by simply "plugging in" point estimates.

We begin by constructing Bayesian prediction intervals (analogous to confidence intervals in classical statistics) for each customer's SOW. By applying equations (16), (17), and (18) to individual draws from the Gibbs Sampler, we are able to sample from the posterior distribution of each customer's SOW and thus determine the posterior probability that each customer meets the loyalty threshold. These probabilities can be used as the basis for classifying customer loyalty.

Figure 1 plots the profits that a retailer would realize from marketing offers that target loyal customers. The plots are based on estimates from the full model specifications using customer information at each of the four grocery retailers. The expected payoff for correctly selecting a loyal customer is shown on the y-axis. This payoff is presented as a multiple of the cost of making an offer. For example, assume that the materials, printing, postage, and handling of mailing a marketing offer costs one dollar. A payoff multiple of two implies that the firm expects an incremental profit of two dollars if the offer is mailed to a correctly identified loyal customer rather than a non-loyal customer. The figure shows a range of payoff multiples from 1.5 to 4 in increments of 0.5. The x-axis shows the Bayesian prediction interval for customer SOW from 0.1 to 0.9 in increments of 0.1. For each level of the Bayesian prediction interval, we compute SOW for all customers and classify them as either loyal or non-loyal based on the 50% SOW threshold. Profit per hundred customers is plotted on the z-axis. It is computed by multiplying the number of correctly classified loyal customers by the payoff multiple, then subtracting the number of offers made (both correctly and incorrectly classified loyals) multiplied by the cost of making an offer. For simplicity, we assume the cost of making an offer to be one dollar.[13]

< Insert Figure 2 about here >

Clearly, profits increase monotonically with the payoff multiple, so the higher the payoff of selecting a loyal customer for the offer, the greater the total profit. On the other hand, profits neither increase nor decrease monotonically with the Bayesian prediction interval used for classification. However, we observe an interesting interaction across the models for Grocery 1, 3, and 4. If the payoff of correctly selecting a loyal customer to receive the offer is low (e.g., 1.5 times cost), it is most profitable to classify customers as loyal only if the probability is very high. This ensures that few customers will be classified as loyal who actually are not. As the payoff of correctly selecting a loyal customer increases, it becomes profitable to make offers to customers who have a lower probability of being loyal. If the payoff is very high (e.g., four times cost) it is profitable to make offers to customers even if the probability that they are loyal is fairly low. Note that this pattern is not observed in the model for Grocery 2. We conjecture that this may be due to the high baseline spending at that retailer (see Table 1).

LIMITATIONS AND FUTURE RESEARCH

In this paper, we have presented an approach to predicting customer SOW using data available to most grocery retailers. We have shown that this approach explains most of the variation in SOW across customers, and offers a high level of predictive accuracy in classifying customer loyalty. Of the three types of information available from card programs, geographic variables appear to offer the most explanation of consumer spending across retailers, while demographics explain the least. By the same token, SOW prediction is critically dependent on the retailer gathering customers' purchase histories so that their spending at its own stores can be observed.

[13] The cost of the offer must exceed the expected payoff of an offer made to a non-loyal customer to ensure that the firm has no incentive to make offers to those outside the targeted segment.

Limitations

One limitation of our approach is that multi-outlet panel data is gathered primarily in metropolitan markets. At present, these panels do not cover rural areas, though the main panel data providers are currently increasing the size of their panels. In addition, panel coverage is relatively sparse outside of the US, making our approach more difficult to implement in other countries. The proposed model would have to be modified to accommodate metropolitan markets with dense urban areas in which consumers travel to the store via public transportation rather than automobiles. In these markets, different parameters for the geographic variables would have to be specified for dense urban areas and areas where consumers drive to the store. Our system of spending models includes only larger competitors in the market, not independent "mom and pop" stores. If smaller retailers are prevalent, SOW predictions could be biased upward. Finally, in order to implement the proposed approach, retailers would need analytical support from service providers with statistical expertise, or would have to have such expertise in house.

Future Research

Additional research opportunities include both extending our approach and using alternative methods. The predictive accuracy of our approach could be improved by optimal selection of shopping behavior variables. The number of possible shopping behavior variables and interactions is unlimited (e.g., units of toilet tissue purchased per person interacted with spending per trip), so variable selection presents an interesting problem. The use of aggregate share models, as opposed to our consumer spending models, to predict SOW also represents an interesting research opportunity. In particular, effectively incorporating information from a shopper loyalty card program in such model provides an interesting challenge.

Table 1
Descriptive Statistics for Retailers

Retailer	N	Spending	Penetration	Store Visits	Travel Time (min)	Retail Concentration*
Grocery 1	1790	$79	0.472	2.6	10.4	15.2
Grocery 2	1790	$184	0.785	7.1	4.9	11.4
Grocery 3	1790	$145	0.570	3.7	8.7	14.9
Grocery 4	1790	$56	0.478	2.4	8.8	13.0
Wal-Mart Supercenter	1790	$122	0.617	4.0	21.2	12.9
Wal-Mart Discount	1790	$30	0.343	1.7	16.8	13.7
Sam's Club	1790	$33	0.298	0.9	20.4	15.2

* Concentration is the number of nearby stores to which a shopper could travel within five minutes

Table 2
Descriptive Statistics for Households

Demographic Variable	N	Average	Std Dev
Income (x $1,000)	358	55.1	30.3
Family Size	358	2.65	1.15
Head of Household Age	358	51.4	11.4
College Education	358	0.38	0.49
Working Woman	358	0.50	0.46

Table 3
Fit Statistics for Consumer Spending Model

			In Sample			Out-of-Sample
Model	Shopper Loyalty Card Data Source	Parameters	LL	Log-Marginal Density	DIC	LL
Full	Grocery 1	133	-18098	-18983	33942	-15089
Full	Grocery 2	133	-17772	-18657	33335	-14635
Full	Grocery 3	133	-17694	-18531	33287	-14445
Full	Grocery 4	133	-17996	-18850	33809	-14949
No Geographic Variables	Grocery 1	119	-21060	-22005	39742	-17174
No Geographic Variables	Grocery 2	119	-20460	-21406	38585	-16666
No Geographic Variables	Grocery 3	119	-20906	-21825	39528	-16370
No Geographic Variables	Grocery 4	119	-20852	-21827	39380	-16726
No Demographics	Grocery 1	98	-18757	-19321	36053	-14807
No Demographics	Grocery 2	98	-18447	-18988	35448	-14364
No Demographics	Grocery 3	98	-18248	-18760	35120	-14456
No Demographics	Grocery 4	98	-18786	-19349	36126	-14777
No Purchase Histories	N/A	105	-18572	-19196	35474	-15137
Baseline	N/A	42	-28528	-	-	-19706

Table 4
Share-of-Wallet Fit Statistics

Model	Source of Purchase History	Number of Customers*	Mean	Std Dev	MAD
Full	Grocery 1	87	0.181	0.215	0.064
Full	Grocery 2	130	0.340	0.283	0.124
Full	Grocery 3	113	0.274	0.298	0.086
Full	Grocery 4	95	0.139	0.187	0.051
No Geographic Variables	Grocery 1	87	0.181	0.215	0.065
No Geographic Variables	Grocery 2	130	0.340	0.283	0.129
No Geographic Variables	Grocery 3	113	0.274	0.300	0.090
No Geographic Variables	Grocery 4	95	0.139	0.187	0.057
No Demographics	Grocery 1	87	0.181	0.215	0.066
No Demographics	Grocery 2	130	0.340	0.283	0.136
No Demographics	Grocery 3	113	0.274	0.300	0.081
No Demographics	Grocery 4	95	0.139	0.187	0.058
No Purchase Histories	Grocery 1	87	0.181	0.215	0.144
No Purchase Histories	Grocery 2	130	0.340	0.283	0.244
No Purchase Histories	Grocery 3	113	0.274	0.300	0.249
No Purchase Histories	Grocery 4	95	0.139	0.187	0.116

* Number of customers who visited the retailer and so would likely have been in its shopper loyalty card program

Figure 1

Confusion Matrices of Loyalty Prediction

Grocery 1

Actual	Predicted: Yes	Predicted: No	Total
Yes	6	6	12
No	5	70	75
	11	76	

Grocery 2

Actual	Predicted: Yes	Predicted: No	Total
Yes	30	6	36
No	15	79	94
	45	85	

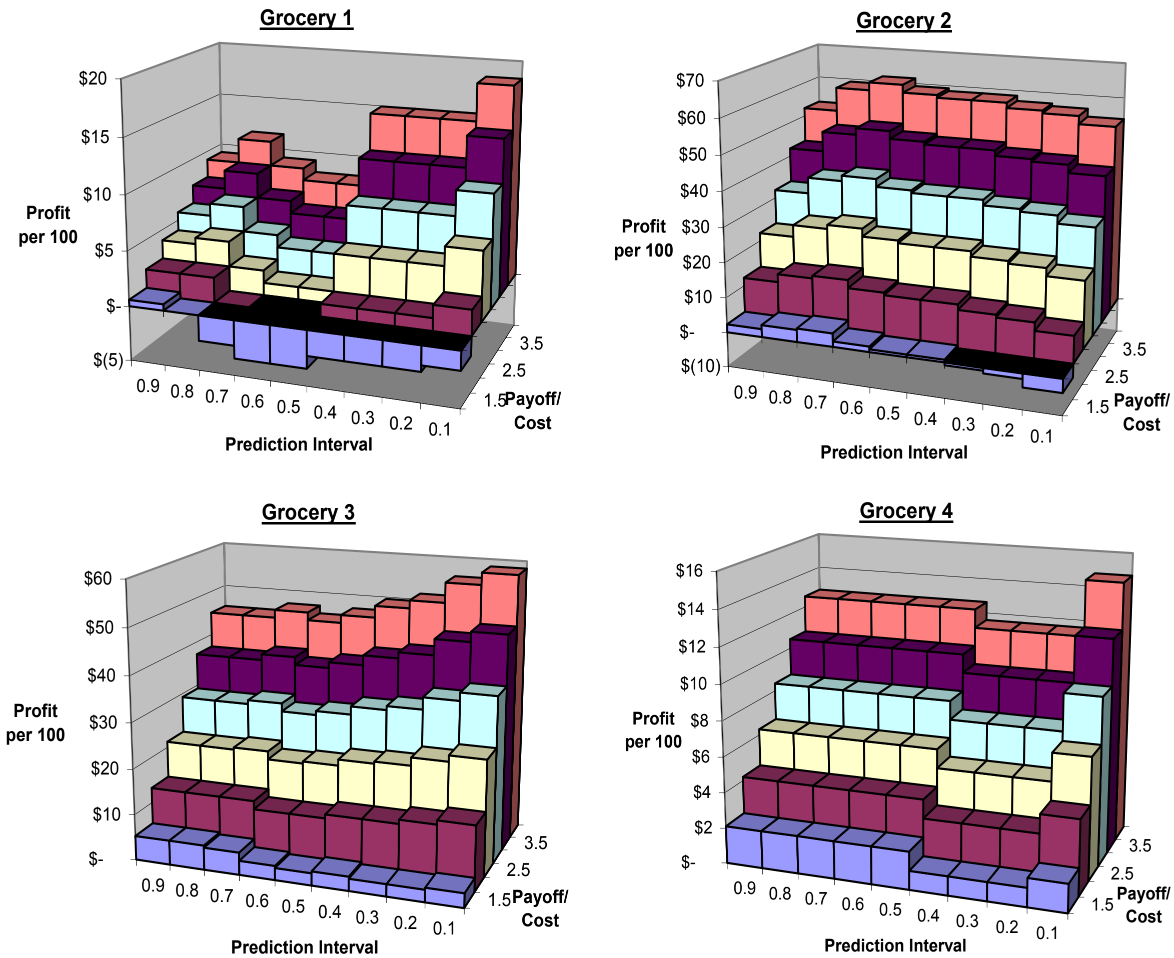
Grocery 3

Actual	Predicted: Yes	Predicted: No	Total
Yes	22	4	26
No	8	79	87
	30	83	

Grocery 4

Actual	Predicted: Yes	Predicted: No	Total
Yes	4	2	6
No	0	89	89
	4	91	

Figure 2
Profitability of Customer Selection for Targeted Marketing Offers



APPENDIX: FULL CONDITIONAL DISTRIBUTIONS

First, we modify our notation. Equation (2) is rewritten in a modified SUR form as follows:

$$\begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \\ \vdots \\ \mathbf{y}_R^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \\ \boldsymbol{\alpha}_R \end{bmatrix} \otimes \left[\mathbf{1}_Q\right] + \begin{bmatrix} X & & & \\ & X & & \\ & & \ddots & \\ & & & X \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_R \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_R \end{bmatrix} \tag{A.1}$$

where: $\mathbf{y}_r^* = \left[y_{1r1}^* \quad y_{1r2}^* \quad \cdots \quad y_{HrQ}^*\right]'$, $\boldsymbol{\alpha}_r = \left[\alpha_{1r} \quad \alpha_{1r} \quad \cdots \quad \alpha_{Hr}\right]'$, $X = \left[\mathbf{1}_H\right] \otimes \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_Q\right]'$,

$\boldsymbol{\varepsilon}_r = \left[\varepsilon_{1r1} \quad \varepsilon_{1r2} \quad \cdots \quad \varepsilon_{HrQ}\right]'$; γ_r is an HQ vector of disturbances such that $E(\gamma_r) = \mathbf{0}$ and $E(\gamma_r\gamma_s\text{N}) =$

$\Phi_{rs}I_{HQ}$, .

For compactness, we rewrite equation (A.1) above:

$$y^* = \alpha + X\theta + \varepsilon \tag{A.2}$$

Similarly, the hierarchical equation (4) is rewritten:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \\ \alpha_R \end{bmatrix} = \begin{bmatrix} Z_1 & & & \\ & Z_2 & & \\ & & \ddots & \\ & & & Z_R \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_R \end{bmatrix} + \begin{bmatrix} \varsigma_1 \\ \varsigma_2 \\ \vdots \\ \varsigma_R \end{bmatrix} \tag{A.3}$$

where: $\alpha_r = \left[\alpha_{1r} \quad \alpha_{2r} \quad \cdots \quad \alpha_{Hr}\right]'$, $Z_r = \left[\mathbf{z}_{1r} \quad \mathbf{z}_{2r} \quad \cdots \quad \mathbf{z}_{Hr}\right]'$, and $\varsigma_r = \left[\varsigma_{1r} \quad \varsigma_{2r} \quad \cdots \quad \varsigma_{Hr}\right]'$; H_r is an H vector of disturbances such that $E(\text{H}_r) = 0$, $E\left(\varsigma_r\varsigma_r'\right) = \omega_r^2 I_H$, and $E\left(\varsigma_r\varsigma_{s\neq r}'\right) = 0$. For compactness, we rewrite equation (A.3) above:

$$\alpha = Z\psi + \varsigma \tag{A.4}$$

The full conditional distributions for our model follow. These conditional distributions also apply to nested models reported in the paper, though some of the parameters in $\mathbf{P}_r$ were restricted to be zero. The word "rest" is used to denote the dataset and all model parameters other than those whose conditional distributions are being specified.

A. $y^*_{hrq} \mid \text{rest} \sim \begin{cases} y_{hrq} & \text{if } y_{hrq} > 0 \\ N_{truncated}\left(\alpha_{hr} + \boldsymbol{\theta}'_{rq} x_q - \boldsymbol{\sigma}_{rs}\Sigma_s^{-1}\left(\mathbf{y}^*_{h,s\neq r,q} - E\left(\mathbf{y}^*_{h,s\neq r,q}\right)\right), \sigma_{rr} - \boldsymbol{\sigma}_{rs}\Sigma_s^{-1}\boldsymbol{\sigma}_{sr}\right) & \text{otherwise} \end{cases}$

where:

$$\Sigma = \begin{bmatrix} \sigma_r^2 & | & \boldsymbol{\sigma}_{rs} \\ -- & + & -- \\ \boldsymbol{\sigma}_{sr} & | & \Sigma_{r\neq s} \end{bmatrix}, \ \mathbf{y}^*_{hq} = \begin{bmatrix} y^*_{hrq} \\ -- \\ \mathbf{y}^*_{h,s\neq r,q} \end{bmatrix}, \text{ and } E\left(\mathbf{y}^*_{h,s\neq r,q}\right) = \left[\alpha_{h2} - \boldsymbol{\theta}'_2\mathbf{x}_q \quad \cdots \quad \alpha_{hR} - \boldsymbol{\theta}'_R\mathbf{x}_q\right]'.$$

As the notation suggests, the $\mathbf{y}^*_{hq}$ vector and Γ matrix are partitioned between the retailer of interest, *r*, and all other retailers, *s* □ *r*. Without loss of generality, we have shown the retailer of interest first.

B. $\boldsymbol{\theta} \mid \text{rest} \sim N\left(O\left(X'\left(\Sigma^{-1} \otimes I_{HQ}\right)\left(\mathbf{y}^* - \boldsymbol{\alpha}\right) + \nu_\theta V_\theta^{-1}\overline{\boldsymbol{\theta}}\right), O\right)$

where $O = \left(X'\left(\Sigma^{-1} \otimes I_{HQ}\right)X + \nu_\theta V_\theta^{-1}\right)^{-1}$

C. $\boldsymbol{\alpha}_h \mid \text{rest} = N\left(P\left(U'\left(\Sigma^{-1} \otimes I_Q\right)\left(\mathbf{y}^*_h - X_h\boldsymbol{\theta}\right) + \Omega^{-1}Z_h\boldsymbol{\psi}\right)\right)$

where U = block diag [$\mathbf{1}_Q$, $\mathbf{1}_Q$, ⌡ $\mathbf{1}_Q$] and $P = \left(U'\left(\Sigma^{-1} \otimes I_Q\right)U + \Omega^{-1}\right)^{-1}$

D. $\boldsymbol{\psi} \mid \text{rest} \sim N\left(Q\left(Z'\left(\Omega^{-1} \otimes I_H\right)\boldsymbol{\alpha} + \nu_\psi V_\psi^{-1}\overline{\boldsymbol{\psi}}\right), Q\right)$

where $Q = \left(Z'\left(\Omega^{-1} \otimes I_H\right)Z + \nu_\psi V_\psi^{-1}\right)^{-1}$

E. $\Sigma^{-1} \mid \text{rest} \sim \text{Wish}\left(HQ + \nu_\Sigma, \left(\nu_\Sigma V_\Sigma + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)^{-1}\right)$

F. $\omega_r^{-2} \mid \text{rest} \sim \chi^2\left(H + \nu_\omega, \left(\nu_\omega \overline{\omega}_r^2 + \varsigma_r\varsigma'_r\right)^{-2}\right)$

REFERENCES

Allenby, Greg M. and Peter E. Rossi. 1999. Marketing models of consumer heterogeneity. *Journal of Econometrics*. **89** 57-78.

Bell, David R., Teck-Hua Ho, and Christopher S. Tang. .1998. Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*. **35**(August) 352-369.

Bowman, Douglas and Das Narayandas. 2001. Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word of mouth behavior. *Journal of Marketing Research*. **38**(August) 281-298.

Chen, Yuxin, and Joel Steckel. 2005. Modeling credit card 'share of wallet': Solving the incomplete information problem." Working paper, Stern School of Business, New York University, New York.

Chib, Siddhartha. 1993. Bayes inference in the Tobit censored regression model. *Journal of Econometrics*. **51** 79-99.

Dowling, Grahame W. 2002. Customer relationship management: In B2C markets, often less is more. *California Management Review*. **44**(Spring) 87-104.

Du, Rex Y., Kamakura, Wagner, and Mela, Carl. 2005. Imputing customer's share of category requirements. Working paper, Fuqua School of Business, Duke University, Durham, NC.

Geweke, John. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Berger, James O., Jose M. Bernado, A. Philip Dawid, and Adrian F. M. Smith (Eds.), *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics*. Clarendon Press, Oxford, UK, 169-193.

Food Marketing Institute. 2004. *Trends in the United States: Consumer Attitudes & the Supermarket, 2003*. Food Marketing Institute, Washington DC.

Fox, Edward J., Alan L. Montgomery and Leonard M. Lodish. 2004. Consumer shopping and spending across retail formats. *Journal of Business*. **77**(2) S25-S60.

Garland, Ron. 2003. Share of wallet's role in customer profitability. *Journal of Financial Services Marketing*. **8**(3) 259-268.

Jarrar, Yasar F., and Andy Neely. .2002. Cross-selling in the financial sector: Customer profitability is the key. *Journal of Targeting, Measurement, and Analysis for Marketing*. **10**(3) 282-296.

Kass, Robert E. and Adrian E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association*. **90** 773-795.

Keiningham, Timothy L., Tiffany Perkins-Munn, Heather Evans. 2003. The impact of customer satisfaction on share-of-wallet in a business-to-business environment. *Journal of Service Research.* **6** 37-50.

Macintosh, Gerrard and Lawrence Lockshin. 1997. Retail relationships and store loyalty: A multi-level perspective. *International Journal of Research in Marketing*. **14**(5) 487-497.

Magi, Anne W. 2003. Share of wallet in retailing: The effects of customer satisfaction, loyalty cards and shopper characteristics. *Journal of Retailing*. **79**(1) 97-106.

Man, Kasing S. 2004. Linear prediction of temporal aggregates under model misspecification. *International Journal of Forecasting*. **20**(4) 659-670.

Miller, Chip E., James Reardon and Denny E. McCorkle. 1999. The effects of competition on retail structure: An examination of intratype, intertype, and intercategory competition. *Journal of Marketing*. **63**(4) 107-120.

Mittal, Vikas. 2004. Research and the bottom line. *Marketing Research*. **16**(Fall) 36-40.

Montgomery, Alan L. 1996. Creating micro-marketing pricing strategies using supermarket scanner data: Technical report. Working paper, Wharton School of the University of Pennsylvania, Philadelphia PA.

Noone, Breffini N., Sheryl E. Kimes, and Leo Renaghan. 2003. Integrating customer relationship management and revenue management: a hotel perspective. *Journal of Revenue and Pricing Management*. **2**(1) 7-22.

Odekerken-Schroder, Gaby, Kristof De Wulf, Hans Kasper, Mirella Kleijnen, Janny Hoekstra, and Harry Commandeur. 2001. The impact of quality on store loyalty: A contingency approach. *Total Quality Management*. **12**(3) 307-322.

Progressive Grocer. 2001.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*. **68**(1) 109-127.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*. **64**(4) 583-639.

Tanner, Martin A. and Wing Hung Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*. **82** 528-540.

Tobin, James. 1958. Estimation of relationships for limited dependent variables. *Econometrica*. **26** 24-36.

Verhoef, Peter. 2003. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*. **67**(4) 30-45.

Wei, Greg C. G. and Tanner, Martin A. 1990. Posterior computations for censored regression data. *Journal of the American Statistical Association*. **85** 829-839.

Willard Bishop Consulting. 1995. *A Guide to Planning Frequent Shopper Programs*. Food Marketing Institute, Washington DC.

Zellner, Arnold. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*. **57** 348-68.