Rex Yuxing Du, Wagner A. Kamakura, & Carl F. Mela

Size and Share of Customer Wallet

Many companies collect substantial information about their interactions with their customers. Yet information about their customers' transactions with competing firms is often sparse or nonexistent. As a result, firms are often compelled to manage customer relationships from an inward view of their customers. However, the empirical analysis in this study indicates that (1) the volume of customers' transactions within a firm has little correlation with the volume of their transactions with the firm's competitors and (2) a small percentage of customers account for a large portion of all the external transactions, suggesting the considerable potential to increase sales if these customers can be correctly identified and incentivized to switch. Thus, the authors argue for a more outward view in customer relationship management and develop a list augmentation–based approach to augment firms' internal records with insights into their customers' relationships with competing firms, including the size of each customer's wallet and the firm's share of it.

With continued advances in information technology, firms are capturing an expanding array of detailed records of their interactions with their individual customers. Such information has been used to generate not only behavioral insights but also important metrics, such as customer lifetime value and customer equity, and it is becoming indispensable in guiding firms' customer relationship management (CRM) initiatives (Rust, Zeithaml, and Lemon 2000). Yet this abundance of internal (within the firm) information is often accompanied by a dearth of information on external (outside the firm) customer activities. As a result, firms are often compelled to manage customer relationships using a view of their customers that is based mostly on internal records. However, such an inward focus could provide misleading measures of a customer's market potential. For example, customers who appear to have high value based on internal records may have modest growth potential if they have only limited requirements that are served by competing firms. In contrast, customers who have a high transactional volume with competing firms may be good targets for growth to the extent that a firm can attract a larger share of their business. Bell and colleagues (2002) indicate that this lack of individual-level, industrywide consumer data is a primary barrier to CRM. In the absence of such information, it may be difficult for customer loyalty programs, cross- and up-selling applications, targeted promotions, and many other marketing efforts to achieve the best return on investment.

Our study attempts to redress this limitation by developing an approach to determine how much business a customer transacts with a focal firm's competitors or, in industry parlance, to estimate the focal firm's share of the customer's total wallet. Such an approach could prove useful in CRM. For example, without information pertaining to a customer's demand from competing firms, firms cannot discern a customer with high firm share and small total wallet from a customer with low firm share and large total wallet. Yet the marketing prescriptions for each differ considerably. For the former customer, the marketing prescription might be to generate new primary demands (if possible), and for the latter customer, it might be more appropriate to encourage switching to the firm's existing products.

Overview of Our Approach

A way to address the problem of not knowing how much business a customer does with the competition is through a procedure commonly known as "list augmentation" or "database augmentation," which overlays data obtained from customer surveys or secondary sources with existing databases (e.g., Crosby, Johnson, and Quinn 2002; Kamakura and Wedel 2003; Kamakura et al. 2003; Wyner 2001). A typical list augmentation exercise involves the following steps: First, the focal firm (often anonymously) surveys a random sample of its customers, collecting information that is not available from the firm's internal database. Second, the information from the survey is linked to information already stored in the internal database (e.g., transaction history, demographics) to form a sample with a complete set of records. Third, from this sample, predictive models that leverage the correlation patterns between the survey results and the internal data can be developed. Finally, the best-performing model is applied to the remainder of the customer database to produce individual-level estimates of the survey results. Compared with the costs of developing and maintaining the customer database, the costs of implementing list augmentation are fairly small.

Central to this general framework is the development of an effective predictive model. In this study, by augmenting the firm's internal database with survey information on a

To read and contribute to reader and author dialogue on JM, visit http://www.marketingpower.com/jmblog.

sample of customers' purchases from the firm's competitors, we present three models for estimating a customer's total purchases in a category and how large a share of these purchases comes from the focal firm. To our knowledge, this is the first study in the marketing literature to use list augmentation to impute wallet size and share of wallet.

We apply our models to a proprietary data set provided by a major U.S. bank. The data set contains information for more than 34,000 customers on their uses of ten categories of financial products that both the bank and its competitors offer. We first calibrate the models on a subsample with information on account balances both inside and outside the bank. We then use the calibrated models to predict total and share of requirements in each category for a validation sample using only inside balances (the outside balances provide the basis for validation), thus emulating the application of the models to the rest of the customer database, in which information about outside balances is unavailable.

Key Findings

One of our three models stands out in terms of its predictive performance and managerial interpretability. It correctly predicts 72% of the time whether a customer uses a product offered by the focal bank's competitors, and it offers the most accurate estimates of total and share of category requirements. Furthermore, it yields insights into customers' share decisions that can be used in guiding future CRM initiatives. We highlight several findings, some of which we conjecture may be idiosyncratic to the bank under study (e.g., Finding 3), whereas others may be generalizable across firms and industries (e.g., Findings 1, 2, and 4).

1. Longer relationships are not necessarily associated with larger share of wallet; the correlation between customer tenure and share of requirements is neutral or negative in four of the ten categories we analyze. This is consistent with the argument that relationship duration and customer share should be considered two separate dimensions of customer relationship (Reinartz and Kumar 2003).
2. Customers with high share in one category also tend to have high share in another category, indicating that customers' share decisions are positively correlated and, thus, the potential for positive externalities across categories.
3. Customers' share and total purchase decisions are sometimes negatively correlated, suggesting that for some categories, customers with small shares within the focal firm tend to transact a large volume outside it. These customers might represent significant opportunities for volume growth to the extent that the focal firm can induce them to switch.
4. Customers with higher incomes tend to balance share of requirements across firms. This may suggest either that the focal firm is not serving such customers well or that customers with higher incomes have incentives to allocate business across firms.

To investigate the managerial value of our proposed approach further, we conduct a series of targeting simulations and find that substantial lifts in targeting efficiency can be obtained by using estimated total and share of wallet. For example, 13% of customers in the validation sample are identified as high-potential customers because their estimated total wallet is in the top quintile and their estimated share of wallet at the focal firm is below average. These customers account for 53% of the validation sample's financial requirements that are fulfilled outside the focal bank, suggesting considerable potential for increasing revenue to the extent that the focal firm can induce them to switch.

We proceed as follows: The next section discusses in more detail the value of share of category requirements in managing relationships with individual customers. We then present the three models for estimating total and share of category requirements. This is followed by our empirical illustration, in which we present the managerial problem and data, the models' predictive performances and parameter estimates, and the results from several customer-targeting simulations. Finally, we summarize and discuss directions for further research.

Share of Category Requirements as an Outward-Looking Relationship Metric

Customer relationship management efforts are often targeted toward a firm's best customers, defined as those who contribute the most to the firm's bottom line. Such strategies are effective when firms want to strengthen relationships with "high-value" customers and mitigate the likelihood that these customers will defect to competing firms. However, it could be myopic to target such customers for relationship development because profitability measures are blind to the relationships a customer might maintain with competitors, and there may be little correlation between the customer's growth potential and his or her current or prior net contributions. Such low correlation may predominate in industries in which customers maintain concurrent relationships with multiple providers. Yet the volume of a customer's business with the competition represents an important source of his or her potential profitability. Indeed, several researchers have proposed or shown the link between customer share and profitability. For example, Garland (2004) examines the role of share of wallet in predicting customer profitability, finding that it is the single relationship-based measure with the most impact on customer contribution. Bowman and Narayandas (2004) and Keiningham, Perkins-Munn, and colleagues (2005) propose that share of wallet mediates the relationship between satisfaction and profits, an effect that Bowman and Narayandas (2004) empirically confirm. Reinartz, Thomas, and Kumar (2005) find that share of wallet positively affects customer tenure and profitability. Zeithaml (2000) proposes a conceptual model in which increased share of wallet is one of four factors that mediate the effect of customer retention on firm profits.

Because scant empirical evidence exists regarding the degree to which a customer's purchases inside and outside a firm correlate, we exemplify this issue using information we obtained from a large sample of customers of a major U.S. bank. For this sample, the bank knows each customer's financial portfolio with both the bank and its competitors. We find that 81% of the sample's internal assets (i.e., the sum of financial assets the sample customers have deposited in the bank) come from customers who are the top 20% in terms of internal asset (this is also known as the

"80/20" rule). Yet these same customers account for only 34% of the sample's external assets (i.e., the sum of financial assets the sample customers have deposited with the bank's competitors). Moreover, the bank's high-volume customers are often not the customers with the greatest growth potential; the correlation between internal assets and external assets is only .13. The pattern is more striking for debt products (e.g., credit card, personal loan, mortgage); 96% of the sample's debts with the focal bank come from customers who are the bank's top 20% clientele in terms of internal debt; however, these customers account for only 20% of the sample's external debts. The correlation between internal debts and external debts is −.04. In short, we find little correlation between a customer's internal and external requirements. Thus, any targeted CRM efforts predicated solely on internal information would miss many high-potential customers who have large sums of assets and debts outside the bank. This suggests that it is especially desirable to consider the bank's share of a customer's total wallet in gauging the customer's potential for growth.

Share of Category Requirements and Share of Wallet

To elaborate on the notion of customer share, we distinguish between share of category requirements and share of wallet. We define share of category requirements as the ratio of (1) a customer's requirements for a particular category of products from a focal supplier to (2) the customer's total requirements for products from all suppliers in the category (i.e., total category requirements). For a firm that offers multiple categories of products to its customers, we define the firm's share of wallet of a customer as the share of total requirements across all the product categories the focal firm offers. Thus, we define share of category requirements at the category level and share of wallet as an aggregate measure across all the categories in which the focal firm competes. We use share of category requirements because of its long-standing status in the literature and share of wallet because of its popularity among practitioners.

Share of category requirements has long been used as a metric of brand loyalty in the context of consumer packaged goods (Fader and Schmittlein 1993), and it is becoming an important metric of customer relationship strength under different names, depending on the industry that is using it (Malthouse and Wang 1998). For example, financial services companies call it "share of wallet," the automobile industry calls it "share of garage," the fashion industry calls it "share of closet," and the media industry calls it "share of eyeballs." Several articles in marketing have studied factors that can affect customer share. For example, Bhattacharya and colleagues (1996) explore the relationship between marketing-mix variables and brand-level share of category requirements. Bowman and Narayandas (2001) assess the impact of customer-initiated contacts on share of category requirements. Keiningham, Perkins-Munn, and Evans (2003) analyze the impact of customer satisfaction on share of wallet in a business-to-business environment. Verhoef (2003) investigates the differential effects of relationship perceptions and marketing instruments on customer reten-

tion and customer share. In the financial services area, researchers have examined the relationship between customer characteristics and share of wallet; for example, Baumann, Burton, and Elliott (2005) use survey data to identify customer characteristics that are associated with high share of wallet in retail banking, and Garland and Gendall (2004) use share of wallet as a factor in predicting customer behavior.
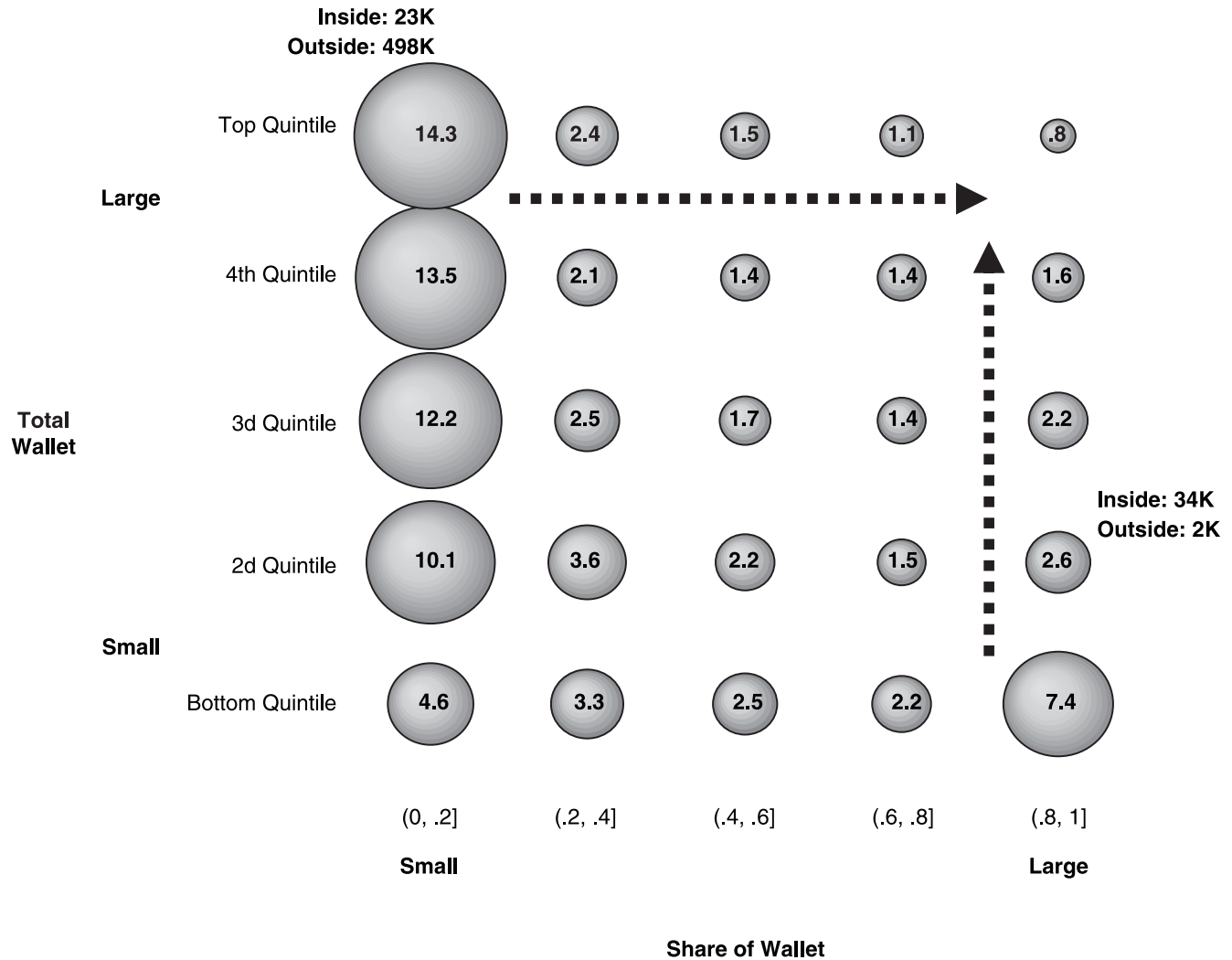
Share of Category Requirements as a Basis for Customer Segmentation

Several researchers have proposed the use of share of wallet as a basis for segmentation. For example, Anderson and Narus (2003) propose a framework for strategically pursuing a customer's business by selecting those with significant incremental share available and superior projected growth in need. Beaujean, Cremers, and Pereira (2005) propose the use of loyalty and share of wallet as bases for customer segmentation. Similarly, Reinartz and Kumar (2003) propose segmenting customers on the basis of customer tenure and share of wallet, and Keiningham, Vavra, and colleagues (2005) propose the combination of share of wallet and customer lifetime value for the same purpose. Although this work is largely theoretical, subsequently, we demonstrate empirically how share of wallet as a segmentation basis can yield actionable strategic insights.

By combining customers' share of wallet with their total wallet and using the bank data (which we describe in more detail subsequently), Figure 1 depicts a 5 × 5 segmentation scheme that offers an outward-looking view of the bank's customer base. The first dimension represents the focal bank's share of a customer's total financial needs (i.e., assets plus debts) divided into the following brackets: (0, 20%], (20%, 40%], (40%, 60%], (60%, 80], and (80%, 100%]. The second dimension is the customer's total financial needs by quintiles. The size of each circle in the figure indicates the percentage of customers who fall into the corresponding segment. Under this scheme, customers in different segments can be viewed as a portfolio of assets with different growth prospects.

Customers in the upper-right-hand corner of Figure 1 are "ideal" from the bank's perspective because they have large requirements for financial products and fulfill most of those requirements with the bank's offerings. However, these customers represent a small fraction of the bank's customer base, as is indicated by the size of their circles. Two growth strategies are indicated: The first is to migrate customers in the upper-left-hand corner toward the right (i.e., gaining a larger share of their business), and the second is to migrate those in the lower-right-hand corner upward (i.e., increasing their category requirements). We conjecture that the former strategy might prove more fruitful in the short run because a customer's category requirements are largely driven by his or her intrinsic needs and are constrained by financial resources. In our data, the two largest cells in the upper-left-hand corner of Figure 1 account for 28% of customers. In terms of short-term growth potential in assets and debts, customers in these cells account for 72%. Thus, we contend that customers in the upper-left-hand corner (i.e., those with a small share of wallet but a large total wal-

FIGURE 1
Segmentation Based on Share of Wallet and Total Wallet

Inside: 23K
Outside: 498K

	(0, .2]	(.2, .4]	(.4, .6]	(.6, .8]	(.8, 1]
Top Quintile	14.3	2.4	1.5	1.1	.8
4th Quintile	13.5	2.1	1.4	1.4	1.6
3d Quintile	12.2	2.5	1.7	1.4	2.2
2d Quintile	10.1	3.6	2.2	1.5	2.6
Bottom Quintile	4.6	3.3	2.5	2.2	7.4

Large

Total Wallet

Small

Inside: 34K
Outside: 2K

Small .... **Large**

Share of Wallet

let) could be the best prospects for marketers seeking near-term growth.

Unfortunately, it is difficult to implement such an outward-looking segmentation and targeting strategy in practice because firms rarely have accurate customer-level measures for share of wallet and total wallet. Without such information, customer segments in the upper-left-hand corner of Figure 1 are indistinguishable from those in the lower right by just looking at requirements fulfilled inside the focal firm. The main purpose of our study is to propose an approach that helps firms estimate these two important measures for their customers.

Share of Category Requirements as a Detector for "Silent Attrition"

Aside from being a segmentation basis, share of wallet can be viewed as an indicator of relationship strength, which can be used in early detection of customer attrition.

Whereas customers who close accounts or move all business to another supplier are clearly defecting, those whose purchases represent a smaller share of their total expenditures are also "defectors" (Reichheld 1996). For example, Coyles and Gokey (2002) find that 5% of customers at a bank close their checking accounts annually, taking with them 3% of the bank's total balances. However, every year, the 35% of customers who reduced their shares with the bank significantly cost the bank 24% of its total balances. This effect was present in all 16 industries they studied (including airlines, banking, and consumer products) and dominant in two-thirds of them. This suggests that partial defection or silent attrition (Malthouse and Wang 1998) caused by decreasing share of wallet can be more serious than attrition, which is detected only when a customer has decided to no longer use the firm's product or service. Accordingly, marketers need to monitor share of wallet on an ongoing basis, and decreasing share of wallet should be

viewed as an early warning signal that a relationship is gradually decaying. Compared with marketing interventions aimed at preventing attrition, marketing efforts that attempt to prevent share-of-wallet losses could be more proactive and, therefore, more effective. Unfortunately, most firms' databases do not contain up-to-date estimates of share of wallet for individual customers.

Share of Category Requirements and Cross-Selling

Since the early 1990s (Kamakura, Ramaswami, and Srivastava 1991), a growing literature has evolved on the topic of cross-selling (Jarrar and Neely 2002; Lau, Chow, and Liu 2004; Lau et al. 2003). This literature is focused on identifying the next product to offer a customer on the basis of the products previously purchased and the patterns of purchase incidence across all customers. More recently, researchers (Kamakura et al. 2003; Li, Sun, and Wilcox 2005) have recognized that customers have multiple relationships and have proposed list augmentation approaches to make cross-selling recommendations that consider the possibility that the customer has already purchased the product elsewhere. To our knowledge, most (if not all) cross-selling models in the marketing literature focus only on the purchase incidence decision. Although cross-selling is an important tool for developing customers, the identification of cross-selling prospects covers only one aspect (purchase incidence) of customer development, overlooking other avenues to enhance the value of a customer. This is particularly true for industries such as banking and retailing, in which customers maintain concurrent relationships with multiple vendors. In these industries, customer relationships can be developed not only by having the customer make purchases in categories in which he or she has not bought before but also by increasing the firm's share of the customer's requirements in categories in which he or she has already made purchases. The list augmentation models we propose and test consider the customer's decision to adopt a product category similar to these cross-selling models, but our models also impute the total volume consumed in the product category and the share of that volume the customer devotes to a particular firm. In other words, we extend the cross-selling models from predicting only the purchase incidence decision to predicting the quantity and incidence decision, thus providing a more informative estimate of a customer's growth potential.

Summary

In light of the foregoing discussion about the advantages and challenges of using share of wallet in CRM, the goal of this article is to develop a predictive model through which a firm can use its internal records, supplemented with a small sample of external records to estimate total and share of requirements in all categories (and, thus, total and share of wallet) for all customers. This enables firms to extend existing customer development initiatives, such as cross-selling, and to benefit from the segmentation, targeting, and attrition detection strategies we discussed previously. Next, we present three models to achieve this aim.

Models for Estimating Total and Share of Category Requirements

The Modeling Task

Consider a firm that offers products in J categories, with transaction records for N customers. The firm knows (1) $Y_{nj}^1 \in Y^1$, how much requirements it serves customer n in category j (superscript 1 denotes the focal firm), and (2) $X_n \in X$, a vector of other customer characteristics (e.g., in our empirical illustration, X consists of customer income and length of relationship with the focal firm). The firm does not have information on $Y_{nj}^0 \in Y^0$, the size of customer n's requirements in category j served by the firm's competitors (superscript 0 denotes outside the focal firm). To learn about customers' outside requirements Y^0, the firm conducts a survey among a random sample, I, of its N customers. The goal of such a survey is to collect two pieces of information in each product category for each customer i in sample I: the self-reported inside requirements Y_{ij}^1 and the self-reported outside requirements Y_{ij}^0. By obtaining self-reported inside requirements, the firm can ensure that they are consistent with their internally recorded counterparts. After data are cleaned to ensure accuracy, the self-reported outside requirements can be linked to records in the firm's customer database. In summary, the firm has complete information (Y^1, Y^0, and X) for only a sample (I) of customers. For the other N – I customers, information on outside requirements, Y^0, is missing. The objective is to develop a model that uses inside requirements, Y^1, and customer characteristics, X, to estimate share of category requirements (S) and total category requirements (T) for these N – I customers, thus augmenting the firm's database with estimates of the customers' relationships with competing firms and, consequently, their potential for growth.

We develop three models to achieve this objective. Model A predicts whether a customer will transact in a category with the focal firm's competitors (i.e., the outside incidence decision) and, if so, the transaction volume (i.e., the outside quantity decision). Model B extends Model A by simultaneously modeling the customers' incidence and quantity decisions both inside and outside the focal firm across multiple product categories, thus explicitly allowing for the possibility that these decisions are correlated. Models A and B can be used to infer the focal firm's share of a customer's category requirements by dividing the customer's inside purchases over the sum of his or her inside and (predicted) outside purchases. In contrast, Model C predicts the share allocation decision directly. It makes three simultaneous predictions: (1) whether a customer will buy in a category, regardless of the vendor (i.e., the category ownership decision), and, if so, (2) the amount to buy in the category (i.e., the total decision) and (3) how large a portion from the focal firm (i.e., the share decision), which can be 0%, 100%, or anything in between.[1] In the empirical sec-

---

[1]Our approach differs from cross-selling or churn analyses insofar as we jointly forecast category ownership, total, and share of requirements in the context of missing data, whereas the churn models that Neslin and colleagues (2006) report, for example, forecast ownership only in the context of complete data.

tion of this study, we calibrate all three models with the same data and compare them on the basis of their predictive performances and managerial interpretability.

Model A: Modeling Incidence and Quantity Outside the Focal Firm

In Model A, we use purchases within the focal firm, along with other customer data available internally, to predict the volume transacted outside the firm. Because the decisions about whether to buy and how much to buy from competitors may be driven by different underlying processes, we propose the following Type II Tobit regression model of incidence and quantity conditional on incidence (Amemiya 1985):

(1)   if $\eta_{ij1}^* > 0$, then $Y_{ij}^0 = \exp(\eta_{ij2}^*)$; else $Y_{ij}^0 = 0$

$$\eta_{ij1}^* = \alpha_{j1} + x_i'\beta_{j1} + \text{Ind}(Y_i^1)'\gamma_{j01} + \ln(Y_i^1)'\gamma_{j11} + \varepsilon_{ij1}$$

$$\eta_{ij2}^* = \alpha_{j2} + x_i'\beta_{j2} + \text{Ind}(Y_i^1)'\gamma_{j02} + \ln(Y_i^1)'\gamma_{j12} + \varepsilon_{ij2},$$

where

- η_{ij1}^* is a latent variable that captures the likelihood of customer i buying in category j from a competing firm. In other words, this latent variable determines the incidence whether customer i has a relationship outside the focal firm in category j;
- Conditional on the customer having a relationship with a competitor, η_{ij2}^* is another latent variable that determines the quantity the customer purchases from the competitor. Because the empirical distributions of conditional quantities are often skewed, they are assumed to be an exponential function of the latent variable and thus enter the likelihood function on a log-transformed scale;
- x_i is a K-element vector of observed characteristics of customer i; $Y_i^1 = (Y_{i1}^1, ..., Y_{iJ}^1)'$; and $\text{Ind}(\cdot)$ is an incidence indicator function, such that $\text{Ind}(Y_{ij}^1) = 1$ if $Y_{ij}^1 > 0$, else $\text{Ind}(Y_{ij}^1) = 0$;
- ε_{ij1} and ε_{ij2} are normally distributed with the covariance

$$\begin{pmatrix} \sigma_{j1}^2 & r_j\sigma_{j1}\sigma_{j2} \\ r_j\sigma_{j1}\sigma_{j2} & \sigma_{j2}^2 \end{pmatrix},$$

and for identification purposes, we assume that $\sigma_{j1} = 1$ (this covariance structure implies that the incidence and quantity decisions might be correlated, which can occur if there are unobserved factors that affect both decisions); and
- α, β, γ, r, and σ are the model parameters, and there are $(2 + 2K + 4J + 2)J$ of them to be estimated.

Model A is relatively simpler than the subsequent ones in that it models the customer's purchase decision in each category independently. Notably, there are likely to be unobserved factors (which therefore cannot be included in Model A as predictors) that affect customers' purchases from both inside (Y_{ij}^1) and outside (Y_{ij}^0) the focal firm. If the impact of these unobserved factors is substantial, it can lead to biased parameter estimates and, thus, to poor predictions of total and share of category requirements. Nonetheless, we still view Model A as a highly practical solution and a strong benchmark for the other two models we propose

because it can be easily estimated as a Type II Tobit regression, and it leads to straightforward data imputations.[2]

Model B: Modeling Incidence and Quantity Inside and Outside the Focal Firm

Model B generalizes Model A by jointly modeling incidence and quantity both inside and outside the focal firm. Modeling decisions inside and outside the firm jointly enable potential correlations between these decisions to be considered. Such correlations might arise from omitted factors that influence inside and outside purchases (e.g., changes in personal status). To the extent that these correlations manifest, inside purchases become informative of outside purchases and therefore can abet the imputation of external demand and customer share. Model B proceeds by assuming that customers make four decisions in each product category: (1) whether to buy from the focal firm; (2) if so, how much; (3) whether to buy from competitors; and (4) if so, how much. Model B allows these decisions to be driven by four different underlying processes that might be correlated not only with one another but also across categories because of the impact of a common set of unobserved customer-specific factors. For example, a recent divorce (for which we do not have data) might lead to reduced purchases from all firms in all categories, suggesting that these decisions could be positively correlated. Formally, we model the incidence and quantity decisions in each category, both inside and outside the focal firm, as follows:

(2)   if $\eta_{ij1}^* > 0$, then $Y_{ij}^1 = \exp(\eta_{ij2}^*)$; else $Y_{ij}^1 = 0$

   if $\eta_{ij3}^* > 0$, then $Y_{ij}^0 = \exp(\eta_{ij4}^*)$; else $Y_{ij}^0 = 0$.

Equation 2 indicates that customer i's incidence and quantity decisions in category j are determined by four latent variables, η_{ij}^*. The top half of Equation 2 implies that customer i will buy a product in category j from the focal firm if η_{ij1}^* is greater than zero. When this happens, the customer's requirements for the focal firm's product are an exponential function of η_{ij2}^*, which implies that the conditional quantities of inside requirements Y_{ij}^1 enter the likelihood function log-transformed. Similarly, the bottom half of Equation 2 states that customer i will buy a product in category j from other firms if η_{ij3}^* is greater than zero, and if so, the customer's requirements served outside the focal firm are an exponential function of η_{ij4}^*.

As indicated previously, we attempt to capture the impact of unobserved customer-specific factors that simultaneously affect customers' incidence and quantity decisions made across product categories and inside and outside

---

[2]With the correlation coefficients (r_j) constrained to zero and a logit link function for incidence, Model A is equivalent to a logistic model for the "whether-to-buy" decision coupled with a log-linear regression model for the "how-much-to-buy" decision conditioned on incidence. The results show that Model A and the logistic/log-linear model with no correlations yield largely identical predictions; thus, we do not discuss this null model further.

the focal firm. Formally, we adopt the following structure on the four latent variables η_{ij1}^*, η_{ij2}^*, η_{ij3}^*, and η_{ij4}^*:

$$(3) \qquad \eta_{ij1}^* = \alpha_{j1} + x_i'\beta_{j1} + z_i'\gamma_{j1} + \varepsilon_{ij1}$$

$$\eta_{ij2}^* = \alpha_{j2} + x_i'\beta_{j2} + z_i'\gamma_{j2} + \varepsilon_{ij2}$$

$$\eta_{ij3}^* = \alpha_{j3} + x_i'\beta_{j3} + z_i'\gamma_{j3} + \varepsilon_{ij3}$$

$$\eta_{ij4}^* = \alpha_{j4} + x_i'\beta_{j4} + z_i'\gamma_{j4} + \varepsilon_{ij4},$$

where

- x_i is defined as previously, a K-element vector of observed characteristics of customer i;
- z_i is a P-element vector that captures unobserved, individual-specific factors that affect the incidence and quantity decisions of customer i; each element of z_i is assumed to be i.i.d. standard normal; ε_{ij1}, ε_{ij2}, ε_{ij3}, and ε_{ij4} are the stochastic components in each decision that are normally distributed with variances σ_{j1}^2, σ_{j2}^2, σ_{j3}^2, and σ_{j4}^2, respectively; for identification purposes, it is assumed that $\sigma_{j1} = \sigma_{j3} = 1$; and
- α, β, γ, and σ are the model parameters; P, the dimensionality of unobserved customer factors, is to be determined empirically; given K, P, and J, $([1 + K + P] \times 4 + 2)J$ parameters need to be estimated.

The foregoing structure provides a parsimonious representation of the correlation pattern between the $4 \times J$ (i.e., the number of decisions in each category times the number of categories) latent variables, η_i^*. Substantial parsimony will be gained when J is large and decisions across these categories are interrelated. As with Model A, a caveat of Model B is that its parameter estimates cannot be directly interpreted to learn about how customers decide share allocations across vendors. Accordingly, in Model C, we model customers' share decisions explicitly.

Model C: Modeling Category Ownership, Total, and the Focal Firm's Share

In Model C, we develop an approach that can be used to predict a customer's category demand and the focal firm's share directly when only inside purchases are recorded. Our motivation for developing such a model is twofold. First, we want to understand how observed customer characteristics affect the total and share-of-category-requirements decisions. This is important because different customer characteristics may be related to a customer's category demand and share allocation in different ways, and marketers who are interested in understanding these relationships can use such insights to devise customer development strategies accordingly. The second reason for modeling total and share of category requirements explicitly is that such a specification aligns itself well with the choice-modeling literature that decomposes consumer purchase decisions into "incidence, quantity, and choice" (Gupta 1988), in which consumers decide whether to buy in a category and, if so, how much and, finally, which brands to choose. In Model C, we attempt to decompose customers' category requirement decisions in a similar way—namely, "ownership, total, and share." Of note, our approach and context is different from that of Gupta (1988) and others insofar as (1) the goal is to impute these decisions with incomplete information, (2) we consider these decisions made across multi-

ple categories, and (3) the choice and share decisions require different treatments.

Consequently, Model C assumes that a customer faces decisions of ownership (whether to own a product category), total (the total category requirements if he or she decides to own), and share (the share of the total requirements, if any, to be served by the focal firm). Formally,

$$(4) \qquad \text{if } \eta_{ij1}^* > 0, \text{ then } T_{ij} = \exp(\eta_{ij2}^*); \text{ else } T_{ij} = 0$$

$$\text{if } \eta_{ij3}^* \leq 0, \text{ then } S_{ij} = 0; \text{ else if } \eta_{ij3}^* \geq 1, \text{ then } S_{ij} = 1;$$

$$\text{else } S_{ij} = \eta_{ij3}^*.$$

Equation 4 posits that customer i's purchase decisions in category j are governed by three latent variables, η_{ij}^*. The top portion of Equation 4 states that the customer will buy a product in the category only if η_{ij1}^* is greater than zero (i.e., category ownership), and when this happens, the customer's total requirements T_{ij} are an exponential function of η_{ij2}^*, which implies that the conditional quantities of total requirements T_{ij} enter the likelihood function log-transformed. The bottom portion of Equation 4 implies that the customer will allocate some of his or her purchases to the focal firm if η_{ij3}^* is greater than zero. When this happens, the customer might have all his or her requirements served by the focal firm (if η_{ij3}^* is greater than or equal to one) or a share of them that is equal to η_{ij3}^*. Thus, we model share of category requirements as a distribution truncated at 0 and 1. Furthermore, we assume that the three latent variables η_{ij1}^*, η_{ij2}^*, and η_{ij3}^* are functions of a common set of factors, observed and unobserved, with a structure as follows:

$$(5) \qquad \eta_{ij1}^* = \alpha_{j1} + x_i'\beta_{j1} + z_i'\gamma_{j1} + \varepsilon_{ij1}$$

$$\eta_{ij2}^* = \alpha_{j2} + x_i'\beta_{j2} + z_i'\gamma_{j2} + \varepsilon_{ij2}$$

$$\eta_{ij3}^* = \alpha_{j3} + x_i'\beta_{j3} + z_i'\gamma_{j3} + \varepsilon_{ij3},$$

where

- x_i and z_i are defined as previously, denoting K observed characteristics and P unobserved factors, respectively, associated with customer i, with each element of z_i assumed to be i.i.d. standard normal; z_i can also be interpreted in the terminology of factor analysis as customer i's scores on P latent factors, in which γ_j are the factor loadings (which we demonstrate how to interpret in the "Results" section); ε_{ij1}, ε_{ij2}, and ε_{ij3} are the stochastic components in each decision, normally distributed with variances σ_{j1}^2, σ_{j2}^2, and σ_{j3}^2, respectively;[3] for identification purposes, we assumed that $\sigma_{j1} = 1$; and
- α, β, γ, and σ are the model parameters; P, the dimensionality of unobserved customer factors, is to be determined empirically; given K, P, and J, $([1 + K + P] \times 3 + 2)J$ parameters need to be estimated.

We denote the variance–covariance matrix of the latent vector η_i^* as $\Lambda(3J \times 3J)$, arising from 3 decisions by J cate-

---

[3]White tests for heteroskedasticity on the log-transformed nonzero balances show that the residuals of the model are not a function of the predicted log total balances, which indicates that the homoskedasticity assumption with regard to the stochastic components conforms to the data.

gories. Thus, for the jth and gth categories (j and g in J) and the hth and lth decisions (ownership, total, or share), the foregoing structure implies that the variance and covariance between η^*_{ijh} and η^*_{igl} can be expressed as follows:

$$(6) \ \mathrm{var}(\eta^*_{ijh}) = \gamma'_{jh}\gamma_{jh} + \sigma^2_{jh} \ \text{ or } \ \mathrm{cov}(\eta^*_{ijh}, \ \eta^*_{igl}) = \gamma'_{jh}\gamma_{gl}.$$

Equation 6 depicts the role of γ, the $3J \times P$ factor loadings, in reducing a large $3J \times 3J$ variance–covariance matrix into a smaller $3J \times P$ factor space as $\Lambda = \gamma \times \gamma' + \Sigma$, where Σ is a diagonal matrix with σ^2_{jh} as entries. As such, the structure depicted in Equation 5 provides a parsimonious representation of the nature and strength of the correlations between $3 \times J$ decisions, namely, ownership, total, and share decisions made by each customer in J product categories. Moreover, the estimates of factor loadings lend themselves to substantive interpretation and graphical display, which we illustrate subsequently in our application. Detailed procedures for estimating this model and for imputation appear in Appendixes A and B, respectively.

Data for Empirical Illustration

Data Collection

A major U.S. bank provided the data we used for our empirical illustration. This bank provided us with information for 34,142 of their customers regarding their balances outside the bank in ten categories ($Y^0_1, \ldots, Y^0_{10}$), including noninterest checking, interest checking, savings, certificates of deposit (CDs), personal managed investments, car loan, personal loan, line of credit, credit card loan, and mortgage. A third-party research firm compiled these data in an ongoing market audit study. Each observation from this audit details when the information was collected, and only households that had at least one account with positive balance with the bank were included to ensure that these customers were active. For the periods during which the market audits were conducted (i.e., from 1999 Q3 to 2002 Q2, with approximately 4500 households each quarter), this information on customer external relationship was aligned with records on the balances these households had inside the bank in the ten categories ($Y^1_1, \ldots, Y^1_{10}$), thus providing a unique data set that includes both total ($T_1, \ldots, T_{10}$) and share of ($S_1, \ldots, S_{10}$) category requirements.

The data set provided by the bank also contains two observed customer characteristics—annual household income and customer tenure—that we use as predictor variables (x_1 and x_2, respectively) in the empirical illustration of our proposed models.[4] Household income is pretax and includes income from both salaries and interest and investment returns. Income-related information is garnered from multiple sources, including Acxiom and customer self-reports (e.g., loan applications), and it is updated peri-

odically through a proprietary "householding" process. Given its skewed nature, household income is log-transformed. Customer tenure is measured in the number of years since the first account was opened at the focal bank.

Because most firms in other industries would not have the benefit of market audit studies conducted by a third-party research supplier, they would need to conduct a survey (often anonymously) on their own with a sample of their customers to obtain data on the volume of business these customers do with competitors. For this sample of customers, with complete information on their relationships with the firm and its competitors, our proposed models can be calibrated and then used to impute total and share of category requirements for all customers. To improve the accuracy of the imputation results, it is desirable to check for self-reporting errors in the survey data. Because firms cannot obtain transactional data from their competitors, it is difficult to check directly the error rates in the sample customers' self-reported data on the business they transact outside the focal firm. An indirect approach is to collect self-reported data on the business the sample customers transact inside the focal firm and to ensure that these data align with internal records. Accordingly, we compared the account balances inside the bank, which we obtained from internal records, with the customers' self-reports, which we obtained through the market audit study. We found little systematic discrepancies; the mean reporting error (market audit reports – internal records) is not significantly different from zero in nine of the ten categories, except for that of savings. Therefore, we presume that the balances outside the bank from the market audit study also exhibit little systematic error.

Study Design

We apportion our data randomly into a calibration sample composed of 23,957 customers and a validation sample of 10,185 customers. This apportionment is intended to reflect the data available to firms in practice, in which they have complete data for only a subset of their customers (analogous to the calibration sample) and incomplete data for the rest of their customer base (analogous to the validation sample).[5] Accordingly, for the calibration sample, we use all the information available, including account balances inside the focal bank (Y^1), household income (x_1), and customer tenure (x_2), as well as account balances outside the focal bank (Y^0). The validation sample mimics the remainder of the customer base for whom the focal bank does not know balances the customers might keep with competitors. We first use the calibration sample to fit the proposed models. Then, in the validation sample, we apply the calibrated models to the internal data (i.e., inside balances, household income, and customer tenure) to impute the outside balances and therefore predict total (T) and share of (S) category requirements. Finally, we evaluate the models' predictive performances in the validation sample by comparing

---

[4]We do not have data for additional behavioral variables, such as the number of branch visits and checks withdrawn. However, the variables we do employ (income and tenure) generalize beyond the banking industry. Adding industry-specific variables as predictors would serve to increase the effectiveness of our approach.

[5]Many firms may not be able to afford a large-scale survey on customers' external transactions. In such cases, because of its efficiency, K-fold cross-validation could be more preferable than the holdout sample cross-validation.

TABLE 1
Structure of the Data Set Used in Empirical Illustration

	Inside the Bank	Outside the Bank
Calibration Sample with 23,957 Customers		
Total and share of category requirements available by combining internal balances with external balances from a market audit study conducted by a third-party research firm	Internal balances in ten accounts Household income Length of relationship with the bank	External balances in ten accounts from the market audit study (Thus, total and share of category requirements observed)
Validation Sample with 10,185 Customers		
Mimicking the remainder of the customer base in which information on total and share of category requirements is unavailable	Internal balances in ten accounts Household income Length of relationship with the bank	Total and share of category requirements to be imputed (Performance evaluated in relation to available data from the market audit study)

the imputed T and S against their observed counterparts. Table 1 summarizes this design.

Summary Statistics

Table 2 reports several summary statistics of the data for both the calibration and the validation samples. Not surprisingly, the two subsamples are similar. Column 2 shows the average category ownership of each type of product. Column 3 shows the average total balances conditional on ownership. Among customers who have positive total balances, Column 4 shows the percentage of customers who have zero balance with the focal bank (thus, share = 0). Column 5 reports the percentage of customers who have positive balances both inside and outside the bank (thus, 0 < share < 1). Approximately 60% of the bank's customers have financial assets inside and outside the bank. Similarly, approximately 50% of the bank's customers have financial debts both inside and outside the bank. Column 6 shows the percentage of customers who are exclusive customers of the focal bank (thus, share = 1). Approximately one in five of the bank's customers have 100% of their requirements for financial assets served by the bank. Fewer than one in ten do so when it comes to fulfilling requirements for financial debts. Together, Columns 4–6 suggest the importance of treating the distribution of share decisions as truncated at both 0 and 1.

Finally, Column 7 shows the bank's average share of requirements in each category. The bank's deposit products (interest and noninterest checking, savings, and CDs) perform best in terms of obtaining a large share of customers' business. Loan products, except for line of credit, have much smaller shares than deposit products. The bank performs most poorly with respect to attracting customers' investment dollars. Taking this information as a whole, the bank has only approximately 20% of its existing customers' total wallet, reflecting high competitive intensity in many of the bank's categories. This competition amplifies the need to use accurate estimates of share of category requirements

to ascertain potential targets across the bank's spectrum of product categories.

The other two variables included in our empirical illustration are the customers' household income and tenure with the focal bank. The mean, median, and standard deviation of income are \$54,566, \$48,086, and \$36,198, respectively, for the calibration sample and \$54,859, \$48,776, and \$35,684, respectively, for the validation sample. The mean, median, and standard deviation of customer tenure (in years) are 9.4, 10, and 3.8, respectively, for the calibration sample and 9.2, 10, and 3.9, respectively, for the validation sample.

Results

Model Comparison

We begin our analysis by comparing the predictive performances of Models A, B, and C in the validation sample. As a basis for comparison, we also consider a log-linear regression of the form $\ln(Y_{ij}^0 + 1) = \alpha_j + x_i'\beta_j + \text{Ind}(Y_i^1)'\gamma_{j0} + \ln(Y_i^1)'\gamma_{j1} + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0, \sigma_j^2)$, because this model can be readily estimated with ordinary least squares.

We consider the ability of each model to predict total and share of category requirements for each separate category, as well as for all assets together, all debts together, and all categories as a whole (i.e., share of wallet and total wallet). We consider three dimensions: outside product ownership, total category requirements (wallet size), and share of category requirements (share of wallet). As Table 3, Panels A and B, indicate, the log-linear model's performance on these dimensions is dominated by the other models. Given that Model A, similar to the log-linear model, can be estimated using off-the-shelf software and dominates the latter, we refrain from further discussion of the log-linear model. We organize our discussion by each dimension.

Predicting outside product ownership. Table 3, Panel A, compares the models' abilities to predict outside product

TABLE 2
Sample Summary Statistics

Product Category / Subsample[c]	Percentage of Customers with Positive Total Category Requirements		Average Total Category Requirements ($)[a]		Percentage of Customers with Share of Category Requirements = 0[a]		Percentage of Customers with 0 < Share of Category Requirements < 1[a]		Percentage of Customers with Share of Category Requirements = 1[a]		Average Share of Category Requirements[a]	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation	Calibration	Validation
Total assets	99.8	99.7	95,191 (226,315)[b]	91,936 (212,861)[d]	20.3	20.1	60.0	59.4	19.7	20.5	17.3	17.3
Noninterest checking	60.4	60.4	5,697 (9,264)	5,810 (10,694)	29.6	29.2	18.8	18.8	51.5	52.0	54.6	55.4
Interest checking	52.0	51.7	10,315 (20,862)	9,420 (17,317)	38.9	38.4	17.3	16.9	43.8	44.6	48.4	49.9
Savings	79.5	79.4	19,912 (56,922)	18,800 (44,414)	43.2	42.8	23.6	24.6	33.2	32.6	36.0	36.2
CDs	19.4	20.2	49,126 (94,765)	48,396 (99,631)	52.7	54.6	13.2	11.4	34.1	34.0	35.1	32.4
Investments	55.8	55.3	108,940 (242,226)	105,996 (231,246)	89.0	89.0	6.5	6.2	4.5	4.9	4.8	4.9
Total debts	87.3	87.2	79,857 (112,534)	78,931 (108,140)	43.4	43.7	48.4	48.3	8.2	8.0	22.4	22.5
Car loan	46.3	46.7	14,106 (12,095)	14,065 (11,986)	80.9	81.6	5.9	5.6	13.3	12.8	14.6	14.2
Personal loan	24.8	24.7	18,746 (29,687)	18,637 (26,938)	81.3	80.5	4.2	4.5	14.5	15.0	16.7	17.2
Line of credit	18.7	18.1	10,521 (24,737)	10,578 (32,399)	50.0	50.7	7.2	7.5	42.8	41.8	39.9	43.3
Credit card	66.3	65.7	4,713 (7,621)	4,682 (6,901)	60.8	60.5	26.0	26.2	13.1	13.3	18.1	17.9
Mortgage	52.1	52.3	102,546 (117,592)	100,760 (111,205)	74.0	73.5	2.0	2.0	24.0	24.5	23.5	23.5

[a] For customers whose total account balances are greater than zero.
[b] Standard deviation is in parentheses.
[c] Number of customers in the calibration sample = 23,957; number of customers in the validation sample = 10,185.
[d] The balances used in estimation are log-transformed to mitigate considerations pertaining to extreme values, and as a result, the standard deviations of the log-transformed balances are smaller than the means, and the means and standard deviations are relatively equal across all product categories.

ownership in terms of the hit rate (i.e., the percentage of true positives and negatives), the percentage of false positives, and the percentage of false negatives.[6] The results show that all three models perform equally well in terms of predicting whether a customer has positive balances outside the bank in a particular category; hit rates, false-positive rates, and false-negative rates average approximately 72~73%, 11~12%, and 16~17%, respectively.

Predicting total category requirements. To compare the models' performance in predicting total category requirements, we calculate the mean absolute deviation (MAD) of the predicted total category requirements.[7] To reflect the variability inherent in our data, we also calculate the MAD for a naive model, in which the predicted total category requirements are simply the account balances inside the bank plus the average account balances outside the bank. Table 3, Panel B (Columns 2–6), reports the naive model's MAD and the percentage improvement (reduction) in MAD relative to the naive model for the various predictive models (which is equal to 1 − MAD_Model/MAD_Naive). The results show that Model C performs much better than Models A and B in predicting the exact sizes of total category requirements (or, equivalently, requirements served outside the focal bank). This is the case for total wallet, total assets, and total debts, as well as for all ten individual product categories.

Predicting share of category requirements. For this comparison, we calculate the MAD of the predicted share of category requirements. By definition, for a category with a zero account balance in the focal bank, share of requirements for that category is either zero (when balance outside the bank in the category is greater than zero) or not defined (when balance outside the bank in the category is also zero); consequently, accuracy in predicting share of requirements for a particular category is relevant only with respect to observations for which there is a positive account balance in that category in the focal bank. Again, as a reference, we use a naive model in which predicted share of category requirements is equal to account balances inside the bank divided by the sum of account balances inside the bank and the average account balances outside the bank. Table 3, Panel B (Columns 7–11), reports the MAD of the naive model, along with the percentage improvement (reduction) in MAD relative to the naive model for the various predictive models. Again, Model C outperforms Models A and B in predicting share of category requirements, which is the case for total wallet, total assets, and total debts, as well as for nine of the ten individual product categories. To our sur-

prise, in eight of the ten categories, the simpler Model A does better than the more sophisticated Model B.

Taking the model comparison results reported in Table 3, Panels A and B, as a whole, we conclude that if the goal is simply to predict whether a customer uses a product offered by the competition (i.e., the cross-selling prediction that Kamakura et al. [2003] propose), any of the three models would be sufficient. Conversely, if the goal is to predict the sizes of total and share of category requirements, Model C is more accurate than the other models. Given that (1) total and share of category requirements are central to assessing customer potential, (2) Model C's parameter estimates can be readily interpreted to gain insights into customers' total and share decisions, and (3) Model C provides the best prediction overall, we believe that Model C has the greatest utility for imputing total and share of category requirements.[8] As such, to conserve space, we subsequently focus on the estimation results for Model C. (Parameter estimates for Models A and B are available on request.)

Estimation Results for Model C

Estimation of Model C from the calibration sample for P = 0–3 latent factors leads us to choose the model with P = 2. The Bayesian information criteria (BICs) for Model C with P = 0, 1, 2, and 3 and for the other two models appear in Table 4. Parameter estimates of the two-factor (P = 2) Model C appear in Table 5, Panel A (which includes the intercepts, α, and the coefficients on customer income and tenure, β) and Panel B (which reports the factor loadings, γ).

Income. Table 5, Panel A, indicates that customer income positively and significantly correlates with both ownership and total decisions across the product categories. Customers with higher incomes are more likely to own assets and debts (e.g., $\beta_{1,income} = .49$ for savings, $p < .01$) and have higher requirements when they own them (e.g., $\beta_{2,income} = .86$ for savings, $p < .01$). Customers with higher incomes also have a significantly smaller share of their requirements for financial products with the bank under study (e.g., $\beta_{3,income} = −.54$ for savings, $p < .01$). We conjecture that there could be many causes for this result. First, with greater requirements for financial products (Asher 2001; Barr 2004) and fewer restrictions (Sharir 1974; Zeithaml 1985), high-income customers stand to gain more by spreading their "nest eggs" across more financial institutions. Second, competition may be more intense for the high-income households' wallets (Bielski 2004), giving them more incentives and opportunities to fulfill their requirements at multiple competing institutions. Third, customer loyalty and satisfaction have been shown to be negatively linked to or moderated by income (Cooil et al. 2007; Homburg and Menon 2003; Korgaonkar, Lund, and Price 1985; Zeithaml 1985). Finally, the bank under study may be comparatively less preferred by its higher-income clientele, for example, because of a weak market position relative to

---

[6]A hit occurs either when the customer has positive external balances and the model predicts that the likelihood of this occurrence will exceed 50% or when the customer has zero balances outside the bank and the model predicts that the likelihood of positive external balances will be below 50%. Depending on the relative costs of misclassification (false positive versus false negative), the threshold (50%) can be adjusted upward or downward.

[7]We also calculated the MAD of the log-transformed predicted total category requirements, and the performance of Model C remained much better than the other two models.

[8]We also estimated Model C with customer income and tenure omitted. The results indicate that most of the model's predictive power arises from internal transactions rather than demographic variables, such as income or tenure.

TABLE 3
Performance Comparisons in the Validation Sample

A: Predicting Outside Product Ownership

	Hit Rate				False-Positive Rate				False-Negative Rate			
	Log-Linear	Model A	Model B	Model C	Log-Linear	Model A	Model B	Model C	Log-Linear	Model A	Model B	Model C
Total Wallet	.37	.73	.72	.72	.63	.11	.11	.12	.00	.16	.17	.16
Total Assets	.38	.73	.72	.72	.62	.11	.11	.12	.00	.16	.17	.16
Noninterest checking	.33	.71	.69	.70	.67	.09	.06	.05	.00	.20	.25	.25
Interest checking	.32	.74	.72	.72	.67	.08	.07	.13	.00	.18	.21	.15
Savings	.54	.66	.64	.64	.46	.19	.22	.23	.00	.15	.14	.13
CDs	.16	.87	.87	.86	.84	.01	.01	.01	.00	.12	.12	.13
Investments	.55	.68	.68	.67	.45	.19	.20	.21	.00	.13	.12	.12
Total Debts	.37	.73	.71	.73	.63	.11	.11	.11	.00	.16	.18	.16
Car loan	.42	.63	.62	.63	.57	.13	.13	.12	.00	.24	.25	.25
Personal loan	.23	.79	.79	.79	.77	.01	.01	.01	.00	.20	.20	.20
Line of credit	.18	.89	.89	.89	.82	.01	.01	.02	.00	.10	.10	.09
Credit card	.57	.62	.61	.61	.43	.27	.29	.28	.00	.11	.10	.11
Mortgage	.46	.71	.65	.71	.54	.13	.12	.13	.00	.16	.23	.16

B: Predicting Total and Share of Category Requirements

	Predicting Total Category Requirements					Predicting Share of Category Requirements				
		Percentage Improvement in MAD					Percentage Improvement in MAD			
Measure	Naive MAD	Log-Linear (%)	Model A (%)	Model B (%)	Model C (%)	Naive MAD	Log-Linear (%)	Model A (%)	Model B (%)	Model C (%)
Total Wallet	125,020	5	6	10	15	.21	6	7	8	12
Total Assets	93,177	4	5	11	18	.39	9	10	11	15
Noninterest checking	2357	10	10	9	17	.34	22	22	17	31
Interest checking	3698	10	11	13	22	.37	24	26	21	37
Savings	13,426	4	4	15	23	.48	21	21	17	37
CDs	11,609	9	10	18	25	.45	16	17	28	36
Investments	76,158	1	2	7	12	.42	4	6	0	15
Total Debts	60,238	8	9	6	14	.26	16	17	0	18
Car loan	7315	7	8	9	11	.33	10	10	-16	15
Personal loan	6222	3	3	6	8	.32	12	13	7	26
Line of Credit	1972	2	-5	-1	5	.39	20	17	25	47
Credit card	3038	7	8	6	8	.37	7	9	8	15
Mortgage	54,322	6	8	3	13	.22	27	30	-116	28

TABLE 4
Fit Measures for the Predictive Models in the Calibration Sample

Number of Factors	Model A Not Applicable	Model B				Model C			
		P = 0	P = 1	P = 2a	P = 3	P = 0	P = 1	P = 2a	P = 3
Number of parameters	460	140	180	220	260	110	140	170	200
BIC	529,695	895,825	878,446	870,232	870,528	838,238	819,865	811,907	811,961

aThe smallest BIC and, thus, the number of factors chosen.

TABLE 5
Model C Parameter Estimates

A: Observed Customer Characteristics (Number of Parameters = 90)

Category	Ownership			Total (Conditional on Ownership)			Share (Conditional on Ownership)		
	α_1 Intercept	β_1 Incomea	β_1 Tenure	α_2 Intercept	β_2 Income	β_2 Tenure	α_3 Intercept	β_3 Income	β_3 Tenure
Noninterest checking	**.78**	**.07**	−.09	**8.01**	**.63**	**.05**	**1.04**	**−.61**	**.56**
Interest checking	**.16**	**1.12**	**.37**	**7.88**	**.76**	**.22**	**.46**	**−.65**	**.75**
Savings	**.91**	**.49**	.06	**8.39**	**.86**	**.38**	**.42**	**−.54**	.27
CDs	**−.96**	**.20**	**.24**	**9.59**	**.24**	**.29**	−.03	**−.76**	**.43**
Investments	**.14**	**.79**	.09	**9.99**	**1.06**	**.49**	**−2.51**	**−.26**	**.41**
Car loan	−.09	**.51**	**−.20**	**9.07**	**.33**	−.02	**−3.64**	**−.39**	−.07
Personal loan	**−.70**	**.28**	**−.20**	**9.09**	**.29**	.02	**−4.75**	**−1.00**	**1.68**
Line of credit	**−.96**	**.41**	**.11**	**7.88**	**.57**	**.25**	**1.18**	**−1.99**	**2.01**
Credit card	**.45**	**.24**	**−.08**	**7.71**	**.41**	.00	**−.31**	**−.14**	.19
Mortgage	.03	**.78**	**.11**	**10.79**	**.51**	**−.15**	**−10.53**	**−.40**	**−.61**

B: Factor Loadings and Variances (Number of Parameters = 80)

Category	Ownership Decision		Total (Conditional on Ownership)			Share (Conditional on Ownership)		
	$\gamma_{1,1}$ Factor 1 Loading	$\gamma_{2,1}$ Factor 2 Loading	$\gamma_{1,2}$ Factor 1 Loading	$\gamma_{2,2}$ Factor 2 Loading	σ_2 SD	$\gamma_{1,3}$ Factor 1 Loading	$\gamma_{2,3}$ Factor 2 Loading	σ_3 SD
Noninterest checking	**−2.76**	**−.30**	−.01	.00	**1.05**	**.16**	**−1.20**	**1.52**
Interest checking	**4.01**	**1.15**	**.67**	**.14**	**1.04**	**.75**	**−1.75**	**1.23**
Savings	.16	**.27**	**.50**	**.12**	**1.48**	**.31**	**−1.45**	**.70**
CDs	**.34**	**.15**	**.40**	**.09**	**1.28**	**.49**	**−2.01**	**2.09**
Investments	.10	**.16**	**.42**	**.13**	**1.53**	**.49**	**−.50**	**1.94**
Car loan	**−.25**	**−.17**	−.03	.00	**.97**	−.11	**.81**	**4.03**
Personal loan	**−.21**	**−.17**	−.02	−.04	**1.19**	**1.36**	**−.49**	**5.68**
Line of credit	−.03	**−.18**	**.09**	**.07**	**1.51**	**.95**	**−4.12**	**3.43**
Credit card	**−.21**	**−.22**	**−.12**	**−.15**	**1.12**	.09	−.02	**1.11**
Mortgage	**−.14**	**−.23**	−.02	**−.09**	**1.36**	**1.97**	**3.43**	**15.66**

aIncome was log-transformed.
Notes: Parameter estimates that are significant at $p < .01$ are in bold.

other nonretail bank providers. In any case, the bank should attend to its lower share of its higher-income customers' wallets and determine whether it should use better-targeted marketing treatments to redress this unfavorable position.

Tenure. The relationship between customer tenure and the customer ownership, total, and share decisions is less clear. Longer tenures are associated with a larger share of customers' checking, CD, investment, personal loan, and line-of-credit balances (as reflected in the positive and significant $\beta_{3,tenure}$). In contrast, there is no significant correlation between customers' tenure and the bank's share of their savings, car loan, and credit card balances (as reflected in

the insignificant $\beta_{3,tenure}$). For customers with longer tenures, the bank has a smaller share of their mortgages ($\beta_{3,tenure} = -.61$, $p < .01$). The lack of a positive relationship between customer tenure and customer share in certain categories (Cooil et al. [2007] also find this for another financial institution) may deserve the bank's close attention because it may imply that lengthening the relationship with customers does not necessarily translate into a larger share of the customers' business in these categories. To pinpoint the dynamics between customer tenure and customer share (Keiningham, Vavra, et al. 2005), however, the bank needs to collect longitudinal data on customer share (instead of using only cross-sectional data).

Factor loadings. Of the 60 estimated factor loadings that appear in Table 5, Panel B, 52 are significant at the $p < .01$ level, indicating that the latent factor structure captures unobserved individual-specific factors that have caused the correlations among ownership, total, and share decisions in the various product categories. To visualize these correlations, Figure 2, Panels A and B, plots the factor loadings from Table 5, Panel B. The rendering of correlation patterns yields managerial insights into customers' ownership, total-wallet, and share-of-wallet decisions among the bank's services. For example, a positive correlation between two ownership decisions means that a customer who is more likely to own one service is also more likely to own another. In Figure 2, Panels A and B, any two vectors pointing in the same (or opposite) direction indicate that decisions these vectors represent are positively (or negatively) correlated; two vectors that are orthogonal to each other imply that the corresponding decisions are independent of each other. We standardized all the factor loadings plotted in the figure so that the length of the vectors and the angle between them directly reflect the strength of correlation between the decisions they represent.

Figure 2, Panel A, captures correlations between the ownership and the total-requirements decisions. Several points are noteworthy:

- There is an assets-versus-debts dimension in terms of the total-requirements decisions. Interest checking, savings, CD, and investment decisions point in one direction, and credit card, mortgage, and personal loan decisions point in the other direction. Thus, all else being equal, customers with more financial assets have fewer debts (except for line of credit). This also provides face validity for the correlation results;

- In terms of the ownership decision, customers who own a noninterest checking and/or a debt account are less likely to own an interest checking account. Ownership of interest checking, savings, CD, and investment accounts are positively correlated, thus reflecting some customers' propensities to save. Similarly, customers with one type of debt are more likely to have other types of debt and are less likely to have financial assets (except for noninterest checking); and

- The same factors that induce customers to assume debt lead them to be deeper in debt (vectors representing debt ownership and total debt point in the same direction). A similar relationship exists for assets.

More germane to targeting customers with high growth potential, the factor loadings plotted in Figure 2, Panel B, depict the correlations between total-requirements and share-of-requirements decisions across various product categories, suggesting the following:

- Typically, share decisions are positively correlated across product categories, except for car loans. Thus, customers who do the bulk of transactions with the bank in one category also tend to do so in others, suggesting the potential for positive externalities across categories;

- In five of the ten categories (i.e., interest checking, savings, CDs, investments, and car loans), the bank has been efficient in targeting in the sense that the bank has larger shares of customers with larger requirements;

- However, for line of credit, total and share of requirements are independent. This implies that the bank might not be targeting high-volume customers to promote its line of credit; and

- For personal loans, credit card, and mortgage, share of category requirements and total category requirements are negatively correlated. This suggests that the focal bank has a low share among high-requirement customers in these three categories. Thus, these customers represent potentially attractive targets.

### Testing the Targeting Efficiency of Our Recommended Approach

To ascertain the efficacy of Model C for targeting customers with large total requirements but small current shares, we assess Model C's ability to identify customers whose total assets (i.e., sum of checking, savings, CDs, and investments) are in the top quintile but the bank's shares are lower than average. Such customers could be denoted as being "platinum" in growth potential under Rust, Zeithaml, and Lemon's (2000) "customer pyramid" framework. We explore Model C's ability to identify these customers by applying the calibrated model to the validation sample and predicting who is in this platinum-potential segment. We then compare these predictions with the actual data.

Although customers we predicted to have platinum potential (i.e., the top quintile in total assets but lower than average shares with the focal bank) constitute only 12% of the validation sample, they account for 37% of the validation sample's assets outside the bank. This leads to a lift of 37%/12% = 3.1 for Model C. Comparable lifts for Models A and B are 2.6 and 2.5, respectively. Using the observed data to identify the platinum-potential segment (i.e., perfect hindsight and maximum lift achievable) yields a lift of 4.7. Together, these results suggest that our list augmentation approach enables firms to target customers with high potential for short-term growth and that Model C performs better than Models A and B at this task.

Similar targeting exercises can also be performed for total debts or on a category-by-category basis. For example, compared with other categories, the focal bank has the smallest share of its customers' investment dollars (see Table 2). Model C suggests that if the bank targets customers whose requirements for investment products are predicted to be in the top quintile and whose share allocated to the bank is predicted to be below average, the bank will be addressing 18% of its customer base that account for only 1% of the investment dollars already inside the bank but

FIGURE 2
Factor Loadings

A: Ownership and Total Decisions
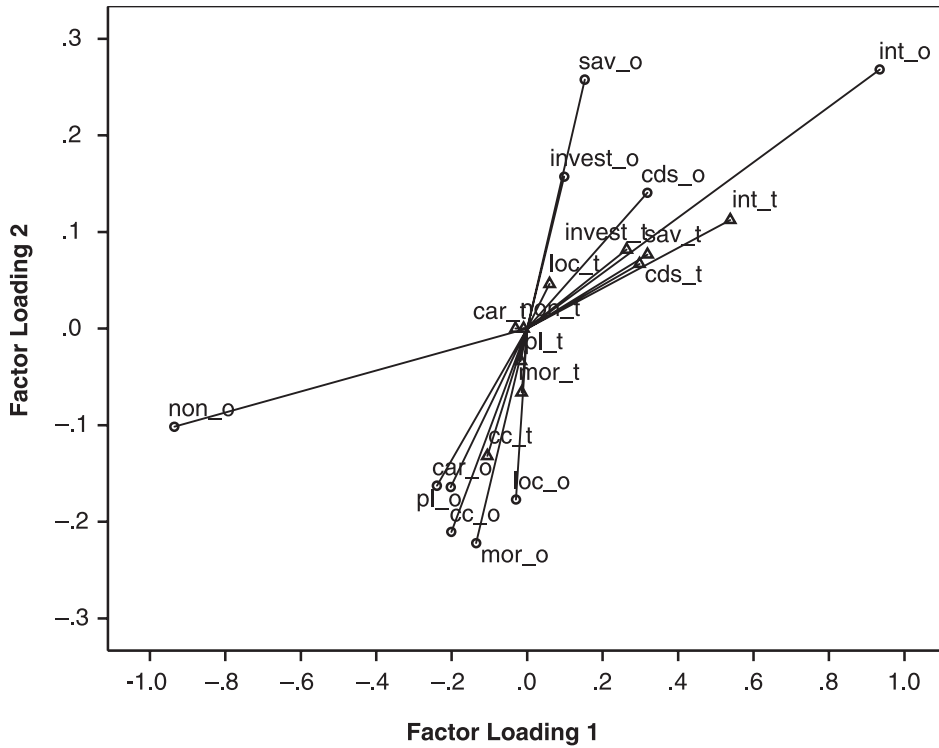
B: Share and Total Decisions
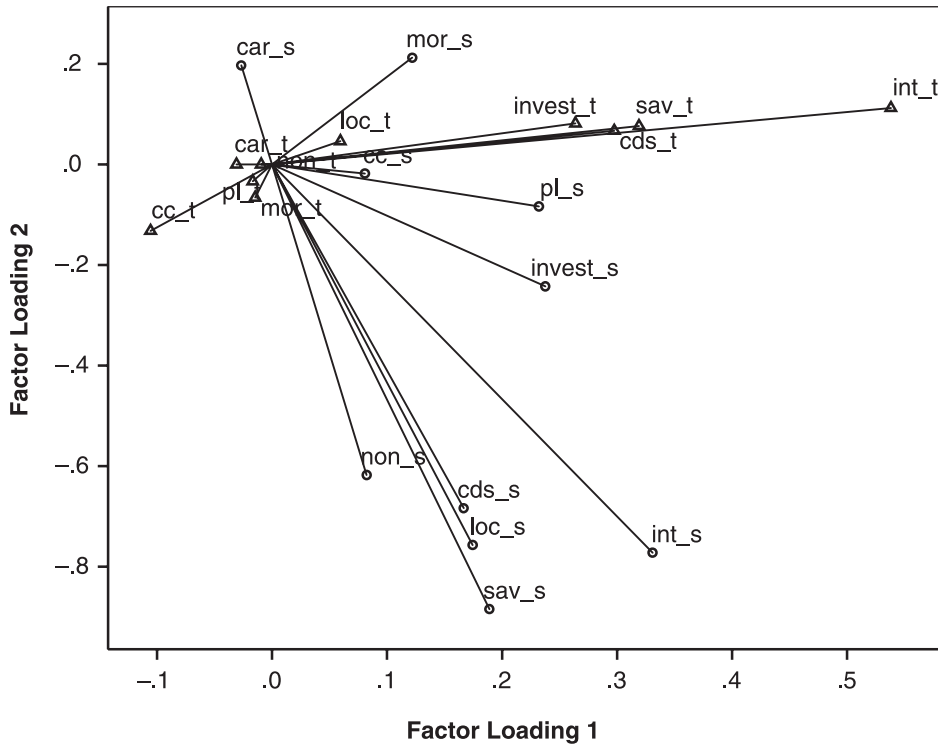
48% of the investment dollars that are currently outside the bank.

Finally, consider the segmentation scheme that Figure 1 illustrates. Using estimated share of wallet and total wallet to segment its customer base, the bank can target the segment on the upper-left-hand corner—namely, those with total wallet in the top quintile and below-average share of wallet. Our approach based on Model C yields 13% of the validation sample customers for targeting. These predicted targets account for 51% of the validation sample's financial requirements that are currently served outside the focal bank. Again, this suggests substantial targeting efficiency (with a lift of 51%/13% = 3.9).

Concluding Remarks

Firms lack individual-level, industrywide customer data because they seldom have information about their customers' relationships with competitors. As a result, many CRM initiatives ignore customers' transactions with competing firms, which we believe reflects a tendency of enterprises taking the new customer-centric paradigm of marketing to an inward-looking extreme. We provide direct evidence that transaction levels inside a firm alone are largely uninformative with respect to a customer's transaction levels outside the firm. This highlights the risks of gauging a customer's potential value by relying solely on profitability recorded in the internal database. Both internal and external data are necessary to distinguish customers with large total market potential and small share from those with small total market potential and large share. Although these two types of customers are indistinguishable on the basis of internally recorded transactions alone, they call for different relationship development strategies.

In general, it is infeasible to obtain transaction records from competing firms, but firms can obtain external relationship data for a small sample of their customers through customer surveys or other secondary sources. We present a list augmentation approach that enables firms to combine internal records with these external data collected for a sample of customers to calibrate a predictive model that can then be used to estimate total and share of requirements for all customers in all categories. In our empirical illustration, we apply three such models to a proprietary data set provided by a major U.S. bank, containing information for more than 34,000 customers' holdings of financial products in ten categories both inside and outside the bank. These data enable us to test the performance of these models in imputing wallet size and share of wallet only on the basis of internal data. We demonstrate that firms can use our approach to segment customers on the basis of share of wallet and total wallet and to discriminate between high-share, small-wallet customers and low-share, large-wallet customers.

One of our three proposed models stands out in several ways: (1) the parsimony of the parameter space, (2) better performance in predicting the sizes of total and share variables and in targeting customers with large wallet but low share, and (3) the ease in interpreting the parameter estimates for behavioral insights (e.g., how customers'

share decisions are correlated across product categories, how these decisions are correlated with total decisions and other customer characteristics). In addition, this model extends extant approaches for imputing data missing because of subsampling (Kamakura and Wedel 2000; Little and Rubin 2002; Schafer 1997) to the context in which multiple unobserved customer decisions—which categories to select, how much to expend in these categories, and how large a share for the focal vendor in each category—are simultaneously imputed from the observed joint outcomes of these decisions (i.e., how much, if any, to expend on the focal vendor's offerings).

Our study also leads to some notable behavioral findings on customers' share decisions. For example, we find that the bank under study has a smaller share of its higher-income customers' wallets. Thus, the bank may want to invest more in its high-income clientele to gain a larger share of their business. In addition, longer customer tenure is not necessarily associated with larger customer share, which runs counter to the conventional wisdom that the longer a customer stays with a company, the more he or she buys from the company. Related to tenure, another avenue for further exploration is to ascertain whether a decreasing share of wallet can be used as an early warning signal to prevent customer attrition, which inevitably entails examining longitudinal share-of-wallet movements. In general, we believe that variations in share of requirements across customers, product categories, and periods should prove to be a fertile ground for response modeling and for evaluating the impacts of different marketing instruments and competitive actions. Insights from these studies can be used to identify the lead indicators and drivers behind such variations, which can then be used to optimize allocation of marketing resources in CRM.

Although we illustrate our approach using data from a major consumer bank, caution should be exercised in extrapolating our specific findings to other institutions and industries. Nonetheless, the modeling framework we developed herein can be readily adapted to any industry in which consumers routinely fulfill their category requirements by purchasing various products/services simultaneously from multiple competing suppliers (e.g., retailing, direct marketing). We believe that share of category requirements can potentially play a more prominent role in improving the understanding and management of customer relationships in a broad range of industries, and our modeling framework could be viewed as a step toward achieving this goal by generating insights into customers' allocation of their business across vendors and by augmenting firms' customer databases with reasonably accurate estimates of total and share of category requirements for each individual customer.

Appendix A
Model Estimation

In Appendix A, we focus on the estimator for Model C. First, let $\lambda_{ijk} \equiv \alpha_{jk} + x_i'\beta_{jk} + z_i'\gamma_{jk}$ for k = 1, 2, or 3. For all possible scenarios of T_{ij} and S_{ij}, we have the following densities conditional on z_i:

When $T_{ij} = 0$, $f(T_{ij}, S_{ij}) = f(\varepsilon_{ij1} \leq -\lambda_{ij1})$, or formally,

(A1) $\qquad f(T_{ij}, S_{ij}|z_i; x_i, \Theta) = \Phi(-\lambda_{ij1}) \equiv \text{Prob}_{ij[1]}(z_i);$

when $T_{ij} > 0$ and $S_{ij} = 1$, $f(T_{ij}, S_{ij}) = f[\varepsilon_{ij1} > -\lambda_{ij1}, \varepsilon_{ij2} = \ln(T_{ij}) - \lambda_{ij2}, \varepsilon_{ij3} \geq 1 - \lambda_{ij3}]$, or

(A2) $\quad f(T_{ij}, S_{ij}|z_i; x_i, \Theta) = \Phi(\lambda_{ij1}) \times \dfrac{1}{\sigma_{j2}} \varphi\left[\dfrac{\ln(T_{ij}) - \lambda_{ij2}}{\sigma_{j2}}\right]$

$$\times \Phi\left(\dfrac{\lambda_{ij3} - 1}{\sigma_{j3}}\right) \equiv \text{Prob}_{ij[2]}(z_i);$$

when $T_{ij} > 0$ and $S_{ij} = 0$, $f(T_{ij}, S_{ij}) = f[\varepsilon_{ij1} > -\lambda_{ij1}, \varepsilon_{ij2} = \ln(T_{ij}) - \lambda_{ij2}, \varepsilon_{ij3} \leq -\lambda_{ij3}]$, or

(A3) $\quad f(T_{ij}, S_{ij}|z_i; x_i, \Theta) = \Phi(\lambda_{ij1}) \times \dfrac{1}{\sigma_{j2}} \varphi\left[\dfrac{\ln(T_{ij}) - \lambda_{ij2}}{\sigma_{j2}}\right]$

$$\times \Phi\left(\dfrac{-\lambda_{ij3}}{\sigma_{j3}}\right) \equiv \text{Prob}_{ij[3]}(z_i); \text{ and}$$

when $T_{ij} > 0$ and $1 > S_{ij} > 0$, $f(T_{ij}, S_{ij}) = f[\varepsilon_{ij1} > -\lambda_{ij1}, \varepsilon_{ij2} = \ln(T_{ij}) - \lambda_{ij2}, \varepsilon_{ij3} = S_{ij} - \lambda_{ij3}]$, or

(A4) $\quad f(T_{ij}, S_{ij}|z_i; x_i, \Theta) = \Phi(\lambda_{ij1}) \times \dfrac{1}{\sigma_{j2}} \varphi\left[\dfrac{\ln(T_{ij}) - \lambda_{ij2}}{\sigma_{j2}}\right]$

$$\times \dfrac{1}{\sigma_{j3}} \varphi\left(\dfrac{S_{ij} - \lambda_{ij3}}{\sigma_{j3}}\right) \equiv \text{Prob}_{ij[4]}(z_i).$$

The terms $\Phi(\cdot)$ and $\varphi(\cdot)$ represent cumulative and probability density functions of the standard normal distribution, respectively. Because z_i is unobservable from a modeler's perspective, we need to integrate it out when deriving the likelihood contribution of individual i:

(A5) $L_i = \displaystyle\int \prod_1 \text{Prob}_{ij[1]}(z_i) \prod_2 \text{Prob}_{ij[2]}(z_i) \prod_3 \text{Prob}_{ij[3]}(z_i)$

$$\prod_4 \text{Prob}_{ij[4]}(z_i) \prod_{p=1}^{P} \varphi(z_{ip}) \, dz_{i1} \cdots \, dz_{iP},$$

where Π_1, Π_2, Π_3, and Π_4 denote the product over four different types of observations across all product categories; an observation belongs to Type 1 if $T_{ij} = 0$, Type 2 if $T_{ij} > 0$ and $S_{ij} = 1$, Type 3 if $T_{ij} > 0$ and $S_{ij} = 0$, and Type 4 if $T_{ij} > 0$ and $1 > S_{ij} > 0$.

For a given P, the dimensionality of the unobserved customer characteristics, we estimate Model C by maximizing the sample likelihood function, which is defined as the product of the individual likelihood functions in Equation A5 across all i's in the calibration subsample of I. We use simulation to evaluate the integrals, which gives rise to a simulated maximum likelihood estimator in which the individual likelihood contributions in Equation A5 can be approximated as

(A6) $\qquad L_i \approx \dfrac{1}{R} \displaystyle\sum_{r=1}^{R} \prod_1 \text{Prob}_{ij[1]}(z_i^r) \prod_2 \text{Prob}_{ij[2]}(z_i^r)$

$$\prod_3 \text{Prob}_{ij[3]}(z_i^r) \prod_4 \text{Prob}_{ij[4]}(z_i^r),$$

where z_i^r is the rth draw from a P-dimensional multivariate standard normal distribution and R is the total number of draws taken. An appealing aspect of the simulated maximum likelihood estimator is that the simulated likelihood function in Equation A6 is twice differentiable, simplifying likelihood maximization with gradient-based search algorithms. We use a unique Halton sequence to simulate each dimension of z_i and assign a different set of draws to each customer. In determining the appropriate R, we adopt the following heuristic: For any given P, we begin with 50P draws and then double the number of draws until no significant improvements can be obtained in estimator efficiency (determined by comparing the standard errors of parameter estimates based on different numbers of draws, as well as the associated likelihoods).

This discussion indicates how our model can be estimated for a given P. To determine the number of factors, we use the BIC. More specifically, we begin with P = 1, then P = 2, and so on, until the resulting BIC stops decreasing.

Appendix B
Imputation Procedure

After calibrating Model C, we are interested in making inferences about the unobserved customer heterogeneities; that is, we impute z_i, the latent factor scores of each individual customer, by combining model parameter estimates and information available for individual i. Depending on the availability of information at the individual level, the formula for imputing z_i varies. We detail the imputation procedures for two scenarios of data availability.

In Scenario 1, both T_{ij} and S_{ij} are observed (or, equivalently, both Y_{ij}^1 and Y_{ij}^0 are observed). Recall that $\Pi_k \text{Prob}_{ij[k]}(z_i)$ for k = 1, 2, 3, or 4, as defined in Appendix A. We determine the marginal density function for T_{ij}, S_{ij}, and z_i as follows:

(B1) $\qquad f(T_{i1}, S_{i1}, ..., T_{iJ}, S_{iJ}, z_i; x_i, \widehat{\Theta})$

$$= \prod_{j=1}^{J} f(T_{ij}, S_{ij}|z_i; x_i, \widehat{\Theta}) \prod_{p=1}^{P} \varphi(z_{ip})$$

$$= \prod_1 \text{Prob}_{ij[1]}(z_i) \prod_2 \text{Prob}_{ij[2]}(z_i)$$

$$\prod_3 \text{Prob}_{ij[3]}(z_i) \prod_4 \text{Prob}_{ij[4]}(z_i)$$

$$\prod_{p=1}^{P} \varphi(z_{ip}) \equiv Q_i(z_i).$$

Because T_{ij} and S_{ij} are observed, this density function can be calculated for any z_i, which enables us to impute cus-

tomer i's latent factor scores by maximizing $Q_i(z_i)$ over z_i. Formally, $\widehat{Z}_i(T_{i1}, S_{i1}, ..., T_{iJ}, S_{iJ}; x_i, \widehat{\Theta}) = \arg\max[Q_i(z_i)]$.

In Scenario 2, only Y_{ij}^1 is observed, whereas T_{ij}, S_{ij}, and Y_{ij}^0 are unobserved. Again, the marginal density function for Y_{ij}^1 and z_i is

(B2) $\quad f(Y_{i1}^1, ..., Y_{iJ}^1, z_i; x_i, \widehat{\Theta}) = \prod_{j=1}^{J} f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$

$$\times \prod_{p=1}^{P} \varphi(Z_{ip}).$$

When $f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$ is available for $\forall j$, $f(Y_{i1}^1, ..., Y_{iJ}^1, z_i; x_i, \widehat{\Theta})$ is defined as

(B3) $\quad \widehat{Z}_i(Y_{i1}^1, ..., Y_{iJ}^1, z_i; x_i, \widehat{\Theta})$

$$= \arg\max\left[f(Y_{i1}^1, ..., Y_{iJ}^1, z_i; x_i, \widehat{\Theta}) \right]$$

$$= \arg\max\left[\prod_{j=1}^{J} f(Y_{ij}^1|z_i; x_i, \widehat{\Theta}) \prod_{p=1}^{P} \varphi(Z_{ip}) \right].$$

Thus, when only Y_{ij}^1 is observed, the central task in imputing z_i is to calculate conditional density $f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$. Depending on whether $Y_{ij}^1 = 0$ or $Y_{ij}^1 > 0$, different formulas are needed to calculate $f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$. When $Y_{ij}^1 = 0$, it can result from two kinds of situations: (1) $T_{ij} = 0$ or (2) $T_{ij} > 0$ and $S_{ij} = 0$. Formally, when $Y_{ij}^1 = 0$,

(B4) $\quad f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$

$$= f(T_{ij} = 0|z_i; x_i, \widehat{\Theta}) + f(T_{ij} > 0, S_{ij} = 0|z_i; x_i, \widehat{\Theta})$$

$$= \Phi(-\lambda_{ij1}) + \Phi(\lambda_{ij1}) \times \Phi\left(\frac{-\lambda_{ij3}}{\sigma_{j3}}\right) \equiv U_{ij[0]}(z_i),$$

where $\lambda_{ijk} \equiv \alpha_{jk} + x_i'\beta_{jk} + z_i'\gamma_{jk}$ for $k = 1, 2,$ or 3.

When $Y_{ij}^1 > 0$, it can also result from two kinds of situations: (1) $T_{ij} = Y_{ij}^1$ and $S_{ij} = 1$ or (2) $T_{ij} \times S_{ij} = Y_{ij}^1$ and $1 > S_{ij} > 0$. Formally, when $Y_{ij}^1 > 0$,

(B5) $\quad f(Y_{ij}^1|z_i; x_i, \widehat{\Theta}) = f(T_{ij} = Y_{ij}^1, S_{ij} = 1|z_i; x_i, \widehat{\Theta})$

$$+ f(T_{ij} \times S_{ij} = Y_{ij}^1, 1 > S_{ij}$$

$$> 0|z_i; x_i, \widehat{\Theta}) =$$

$$f(T_{ij} = Y_{ij}^1, S_{ij} = 1|z_i; x_i, \widehat{\Theta}) + \int_{1 > S > 0} f(T_{ij} = \frac{Y_{ij}^1}{S}|S_{ij}$$

$$= S, z_i; x_i, \widehat{\Theta})f(S_{ij} = S|z_i; x_i, \widehat{\Theta})dS =$$

$$\Phi(\lambda_{ij1}) \times \left\{ \frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln(Y_{ij}^1) - \lambda_{ij2}}{\sigma_{j2}}\right]\Phi\left(\frac{\lambda_{ij3} - 1}{\sigma_{j3}}\right) + \int_{1 > S > 0} \frac{1}{\sigma_{j2}} \right.$$

$$\left. \varphi\left[\frac{\ln\left(\frac{Y_{ij}^1}{S}\right) - \lambda_{ij2}}{\sigma_{j2}}\right]\frac{1}{\sigma_{j3}}\varphi\left(\frac{S - \lambda_{ij3}}{\sigma_{j3}}\right)dS \right\} \equiv U_{ij[1]}(z_i).$$

Combining this, we obtain the following:

(B6) $\quad f(Y_{i1}^1, ..., Y_{iJ}^1, z_i; x_i, \widehat{\Theta}) = \prod_{j=1}^{J} f(Y_{ij}^1|z_i; x_i, \widehat{\Theta})$

$$\prod_{p=1}^{P} \varphi(Z_{ip}) = \prod_0 U_{ij[0]}(z_i) \prod_1 U_{ij[1]}(z_i) \prod_{p=1}^{P} \varphi(Z_{ip}) \equiv V_i(Z_i),$$

where Π_0 and Π_1 denote the product over two different types of observed Y_{ij}^1 across all product categories; an observation belongs to Type I if $Y_{ij}^1 = 0$ and to Type II if $Y_{ij}^1 > 0$. To impute the latent factor scores when only Y_{ij}^1 is observed, we can maximize $V_i(z_i)$ over z_i. Formally, $\widehat{Z}_i(Y_{i1}^1, ..., Y_{iJ}^1; x_i, \widehat{\Theta}) = \arg\max[V_i(z_i)]$.

We now discuss the prediction of T_{ij} and S_{ij}. Given $\widehat{\Theta}$, the estimated model parameters, and $\widehat{z}_i$, the imputed latent factor scores, we can predict T_{ij}, customer i's total category requirements, and S_{ij}, shares of category requirements that are served by the focal firm, conditional on (1) observed Y_{ij}^1, levels of category requirements that are served by the focal firm, and (2) x_i, observable customer characteristics. (Instead of plugging in $\widehat{z}_i$, an alternative is to integrate z_i out, which in practice leads to predictions that are of no significant differences but could take significantly more time to implement when the dimensionality of z_i is high.) The rest of this section demonstrates the formulas for making these predictions.

For categories in which $Y_{ij}^1 = 0$, we can predict, for example, the expected value of T_{ij} conditional on $T_{ij} > 0$,

(B7) $\quad E(T_{ij}|T_{ij} > 0, Y_{ij}^1 = 0, \widehat{z}_i; x_i, \widehat{\Theta}) = \exp\left(\lambda_{ij2} + \frac{\sigma_{j2}^2}{2}\right),$

and the expected value of S_{ij} conditional on $T_{ij} > 0$,

(B8) $\quad E(S_{ij}|T_{ij} > 0, Y_{ij}^1 = 0, \widehat{z}_i; x_i, \widehat{\Theta}) = 0.$

Similarly, for categories in which $Y_{ij}^1 > 0$, we can make the foregoing as follows: The expected value of T_{ij} conditional on $T_{ij} > 0$ is

(B9) $\quad E(T_{ij}|T_{ij} > 0, Y_{ij}^1 = Y > 0, \widehat{z}_i; x_i, \widehat{\Theta})$

$$= \left\{ \left(\frac{Y_{ij}^1}{1}\right)\frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(Y_{ij}^1\right) - \lambda_{ij2}}{\sigma_{j2}}\right]\Phi\left(\frac{\lambda_{ij3} - 1}{\sigma_{j3}}\right) \right.$$

$$+ \int_{1 > S > 0} \frac{Y_{ij}^1}{S}\frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(\frac{Y_{ij}^1}{S}\right) - \lambda_{ij2}}{\sigma_{j2}}\right]\frac{1}{\sigma_{j3}}$$

$$\left. \varphi\left(\frac{S - \lambda_{ij3}}{\sigma_{j3}}\right)dS \right\} \Big/ \left\{ \frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(Y_{ij}^1\right) - \lambda_{ij2}}{\sigma_{j2}}\right] \right.$$

$$\Phi\left(\frac{\lambda_{ij3}-1}{\sigma_{j3}}\right) + \int_{1>S>0} \frac{1}{\sigma_{j2}}$$

$$\varphi\left[\frac{\ln\left(\frac{Y_{ij}^1}{S}\right)-\lambda_{ij2}}{\sigma_{j2}}\right]\frac{1}{\sigma_{j3}}\varphi\left(\frac{S-\lambda_{ij3}}{\sigma_{j3}}\right)dS\Bigg\},$$

and the expected value of S_{ij} conditional on $T_{ij} > 0$ is

(B10) $E(S_{ij}|T_{ij} > 0,\ Y_{ij}^1 = Y > 0,\ \hat{z}_i;\ x_i,\ \hat{\Theta})$

$$= \left\{1 \times \frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(Y_{ij}^1\right)-\lambda_{ij2}}{\sigma_{j2}}\right]\Phi\left(\frac{\lambda_{ij3}-1}{\sigma_{j3}}\right)\right.$$

$$+ \int_{1>S>0} S\times\frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(\frac{Y_{ij}^1}{S}\right)-\lambda_{ij2}}{\sigma_{j2}}\right]\frac{1}{\sigma_{j3}}$$

$$\varphi\left(\frac{S-\lambda_{ij3}}{\sigma_{j3}}\right)dS\Bigg\} \Bigg/ \left\{\frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(Y_{ij}^1\right)-\lambda_{ij2}}{\sigma_{j2}}\right]\right.$$

$$\Phi\left(\frac{\lambda_{ij3}-1}{\sigma_{j3}}\right) + \int_{1>S>0}\frac{1}{\sigma_{j2}}\varphi\left[\frac{\ln\left(\frac{Y_{ij}^1}{S}\right)-\lambda_{ij2}}{\sigma_{j2}}\right]$$

$$\frac{1}{\sigma_{j3}}\varphi\left(\frac{S-\lambda_{ij3}}{\sigma_{j3}}\right)dS\Bigg\}.$$

REFERENCES

Amemiya, Takeshi (1985), *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Anderson, James C. and James A. Narus (2003), "Selectively Pursuing More of Your Customer's Business," *Sloan Management Review*, 44 (3), 42–49.

Asher, Joseph (2001), "'Wealth Management' Moves Center Stage," *ABA Banking Journal*, 93 (April), 41–46.

Barr, Michael S. (2004), "Banking the Poor," *Yale Journal on Regulation*, 21 (1), 121–237.

Baumann, Chris, Suzan Burton, and Greg Elliott (2005), "Determinants of Customer Loyalty and Share of Wallet in Retail Banking," *Journal of Financial Services Marketing*, 9 (3), 231–48.

Beaujean, Marc, Vincent Cremers, and Francisco Pedro Goncalves Pereira (2005), "How Europe's Banks Can Profit from Loyal Customers," *McKinsey Quarterly*, (November), 16–19.

Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002), "Seven Barriers to Customer Equity Management," *Journal of Service Research*, 5 (August), 77–86.

Bhattacharya, C.B., Peter S. Fader, Leonard M. Lodish, and Wayne S. DeSarbo (1996), "The Relationship Between Marketing Mix and Share of Category Requirements," *Marketing Letters*, 7 (1), 5–18.

Bielski, Lauren (2004), "Slow Buildup for 'Mass Affluent' Success," *ABA Banking Journal*, 96 (July), 61.

Bowman, Douglas and Das Narayandas (2001), "Managing Customer-Initiated Contact with Manufacturers: The Impact on Share of Category Requirements and Word-of-Mouth Behavior," *Journal of Marketing Research*, 38 (August), 281–97.

——— and ——— (2004), "Linking Customer Management Effort to Customer Profitability in Business Markets," *Journal of Marketing Research*, 41 (November), 433–47.

Cooil, Bruce, Timothy L. Keiningham, Lerzan Aksoy, and Michael Hsu (2007), "A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics," *Journal of Marketing*, 71 (January), 67–83.

Coyles, Stephanie and Timothy C. Gokey (2002), "Customer Retention Is Not Enough," *McKinsey Quarterly*, (2), 81–89.

Crosby, Lawrence A., Sheree L. Johnson, and Richard Quinn (2002), "Is Survey Research Dead?" *Marketing Management*, 11 (May–June), 24–29.

Fader, Peter S. and David C. Schmittlein (1993), "Excess Behavioral Loyalty for High-Share Brands: Deviations from the Dirichlet Model for Repeat Purchasing," *Journal of Marketing Research*, 30 (November), 478–93.

Garland, Ron (2004), "Share of Wallet's Role in Customer Profitability," *Journal of Financial Services Marketing*, 8 (3), 259–68.

——— and Philip Gendall (2004), "Testing Dick and Basu's Customer Loyalty Model," *Australasian Marketing Journal*, 12 (3), 81–87.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342–55.

Homburg, Christian and Ajay Menon (2003), "Relationship Characteristics as Moderators of the Satisfaction-Loyalty Link: Findings in a Business-to-Business Context," *Journal of Business-to-Business Marketing*, 10 (3), 35–62.

Jarrar, Yasar F. and Andy Neely (2002), "Cross-Selling in the Financial Sector: Customer Profitability Is Key," *Journal of Targeting, Measurement and Analysis for Marketing*, 10 (3), 282–96.

Kamakura, Wagner A., S.N. Ramaswami, and R.K. Srivastava (1991), "Applying Latent Trait Analysis in the Evaluation Process for Cross-Selling of Financial Services," *International Journal of Research in Marketing*, 8 (4), 329–49.

——— and Michel Wedel (2000), "Factor Analysis and Missing Data," *Journal of Marketing Research*, 37 (November), 490–98.

——— and ——— (2003), "List Augmentation with Model Based Multiple Imputation: A Case Study Using a Mixed-Outcome Factor Model," *Statistica Neerlandica*, 57 (1) 46–57.

———, ———, Fernando de Rosa, and Jose Afonso Mazzon (2003), "Cross-Selling Through Database Marketing: A Mixed Data Factor Analyzer for Data Augmentation and Prediction," *International Journal of Research in Marketing*, 20 (1), 45–65.

Keiningham, Timothy L., Tiffany Perkins-Munn, Lerzan Aksoy, and Demitry Estrin (2005), "Does Customer Satisfaction Lead to Profitability? The Mediating Role of Share of Wallet," *Managing Service Quality*, 15 (2), 172–81.

———, ———, and Heather Evans (2003), "The Impact of Customer Satisfaction on Share-of-Wallet in a Business-to-Business Environment," *Journal of Service Research*, 6 (August), 37–50.

———, Terry G. Vavra, Lerzan Aksoy, and Henri Wallard (2005), *Loyalty Myths: Hyped Strategies That Will Put You Out of Business and Proven Tactics That Really Work*. Hoboken, NJ: John Wiley & Sons.

Korgaonkar, Pradeep K., Daulat Lund, and Barbara Price (1985), "A Structural Equations Approach Toward Examination of Store Attitude and Store Patronage Behavior," *Journal of Retailing*, 61 (2), 39–60.

Lau, Kin-Nam, Haily Chow, and Connie Liu (2004), "A Database Approach to Cross Selling in the Banking Industry: Practices, Strategies and Challenges," *Journal of Database Marketing & Customer Strategy Management*, 11 (3), 216–34.

———, Sheila Wong, Margaret Ma, and Connie Liu (2003), "'Next Product to Offer' for Bank Marketers," *Journal of Database Management*, 10 (4), 353–68.

Li, Shibo, Baohong Sun, and Ronald T. Wilcox (2005), "Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research*, 42 (May), 233–39.

Little, Roderick J.A. and Donald B. Rubin (2002), *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons.

Malthouse, Edward C. and Paul Wang (1998), "Database Segmentation Using Share of Customer," *Journal of Database Marketing*, 6 (3), 239–52.

Neslin, Scott A., Sunil Gupta, Wagner Kamakura, Junxiang Lu, and Charlotte H. Mason (2006), "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research*, 43 (May), 204–211.

Reichheld, Frederick F. (1996), *The Loyalty Effect*. Boston: Harvard Business School Press.

Reinartz, Werner J. and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (January), 77–99.

———, Jacquelyn S. Thomas, and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69 (January), 63–79.

Rust, Roland T., Valarie A. Zeithaml, and Katherine N. Lemon (2000), *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy*. New York: The Free Press.

Schafer, Joe L. (1997), *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Sharir, Samuel (Shraier) (1974), "Brand Loyalty and the Household's Cost of Time," *Journal of Business*, 47 (1), 53–55.

Verhoef, Peter (2003), "Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development," *Journal of Marketing*, 67 (October), 30–45.

Wyner, Gordon A. (2001), "When 360 Degrees Is Not Enough," *Marketing Management*, 10 (July–August), 4–5.

Zeithaml, Valarie A. (1985), "The New Demographics and Market Fragmentation," *Journal of Marketing*, 49 (July), 64–75.

——— (2000), "Service Quality, Profitability, and the Economic Worth of Customers: What We Know and What We Need to Learn," *Journal of the Academy of Marketing Science*, 28 (1), 67–85.