RAJKUMAR VENKATESAN, V. KUMAR, and TIMOTHY BOHLING*

This study addresses significant challenges that practitioners face when using customer lifetime value (CLV) for customer selection. First, the authors propose a Bayesian decision theory–based customer selection framework that accommodates the uncertainty inherent in predicting customer behavior. They develop a joint model of purchase timing and quantity that is amenable for selecting customers using CLV. Second, the authors compare performance of the proposed customer selection framework (1) with the current customer selection procedure in the collaborating firm and (2) with different customer-level cost allocation rules that are necessary for computing CLV. The study finds that given a budget constraint, customers selected by means of a Bayesian decision theory–based framework (i.e., using the maximized expected CLV of a customer and the corresponding optimal marketing costs as an estimate of future costs) provide the highest profits. The study provides guidelines for implementation and illustrates how the proposed customer selection framework can aid managers in enhancing marketing productivity and estimating return on marketing actions.

Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection

Marketing practitioners across industries are under increased pressure to measure and maximize the return on marketing investment to improve the value of the firm. This has increased the importance of identifying marketing assets in which to invest and of understanding how the assets provide potential for sustained profits in the long run (Rust, Lemon, and Zeithmal 2004). Customers are considered a critical element of a firm's marketing assets, and the effective management of customer assets is expected to affect firm profits directly (Bolton, Lemon, and Verhoef 2004). In this context, the emphasis has shifted toward measuring the value of customer assets; understanding the impact of marketing expenditures on customer value; and actively using marketing actions, such as contacts through various channels, including salesperson and direct mail, to maximize customer value and, thus, firm value (Webster 1992).

Given an unlimited marketing budget, managers can contact all their customers at every period. Such a strategy minimizes the Type I error of not contacting a customer who could have potentially provided revenue. However, minimizing Type I error also maximizes a so-called Type II error. A Type II error is to contact a customer who is not ready to purchase and is costly in terms of adversely affecting both the bottom line and the top line (Venkatesan and Kumar 2004). When faced with a limited marketing budget, the trade-offs between Type I error and Type II error are highlighted further, and managers are forced to prioritize their communication strategies toward customers who are expected to provide the highest growth in cash flows (i.e.,

*Rajkumar Venkatesan is Associate Professor of Business Administration, Darden Graduate School of Business, University of Virginia, Charlottesville (e-mail: Venkatesanr@darden.virginia.edu). V. Kumar (VK) is ING Chair Professor in Marketing and Executive Director, ING Center for Financial Services, School of Business, University of Connecticut (e-mail: vk@business.uconn.edu). Timothy Bohling is Vice President, Market Intelligence, IBM Americas, Armonk, NY (e-mail: tbohling@us.ibm.com). The authors thank the anonymous *JMR* reviewers for their valuable comments on previous versions of this article. They thank a multinational firm not only for providing access to the data used in this study but also for actively participating in the execution of the study. The study benefited from presentation at numerous places, including the Marketing Science Institute/Yale conference on practitioner–academic collaboration and the IBM corporation. The authors thank Denise Beckmann, Dipak Dey, Gary Lilien, Leigh McAlister, Anil Menon, Scott Neslin, and Nalini Ravishanker for their suggestions on how to improve the contribution of this manuscript. They also thank Renu for the initial copyediting of the submitted manuscript. Peter Lenk served as guest associate editor for this article.

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

ISSN: 0022-2437 (print), 1547-7193 (electronic)

Journal of Marketing Research
Vol. XLIV (November 2007), 579–594

customer selection). Therefore, accurate measurement or estimation of the future value of customers is critical for the success of customer selection in a firm.

Conceptually, customer lifetime value (CLV), which is the net present value (NPV) of long-term cash flows from a customer, is regarded as an appropriate measure for customer selection (Venkatesan and Kumar 2004). To a large extent, empirical evidence regarding the weak correlation between customer profits and behavioral measures of loyalty, such as customer lifetime duration (Reinartz and Kumar 2000), has also contributed to the widespread adoption of CLV. In this study, we address three significant practitioner concerns about the utilization of CLV, which includes determining (1) the optimal level of contacts, (2) the return-on-investment implications of maximizing CLV, and (3) how best to manage the implementation of CLV for customer selection.

UNCERTAINTY IN PREDICTING CUSTOMER BEHAVIOR

Theoretically, CLV models should estimate the value of a customer over the customer's lifetime. However, in many firms, including our collaborating firm, three years is considered a good estimate for the horizon over which the current business environment (e.g., with regard to technology or competition) would not change substantially. Thus, most customer relationship management (CRM) decisions, including customer selection, are made on the basis of CLV estimates over a three-year window, and even then, there is significant uncertainty in predicting customer behavior. Typical sources for this uncertainty are related to the poor or nonexistent information in most CRM databases on (1) customer transactions with the competition, (2) competitor marketing actions targeted at each customer, and (3) customer attitudes. Sometimes, the cost of increased errors can outweigh the benefits of long-term predictions. In this study, therefore, we propose a Bayesian decision theory–based customer selection framework that directly accommodates this uncertainty and reduces the possibility of errors when making long-term predictions.

CUSTOMER-LEVEL COST ALLOCATION RULE

Current practice among both practitioners and academicians is to use a naive status quo cost allocation strategy for calculating forward-looking metrics, such as CLV. In this status quo strategy, the cost of serving the customer in the most recent year is assumed to be the customer costs that would prevail in the future periods over which CLV is calculated. However, this practice is inconsistent with the core assumption for using forward-looking metrics over backward-looking metrics (i.e., past customer profitability is not the best indicator of future customer profitability). Thus, similar to the metrics, the cost allocation rules also need to be forward looking. Two methods are possible for a forward-looking cost allocation rule: (1) using a regression-based estimate for the future costs of serving a customer and (2) using the optimal cost that maximizes the expected value of CLV as the future costs of serving a customer. An optimal cost allocation rule dynamically updates the estimates of future customer costs on the basis of customer responsiveness to historic marketing communication. In addition, marketing practitioners and scholars recognize that the channels of marketing communications and the levels of communication need to be customized to individual customer preferences (Shultz 2003) and that there is scope for substantial improvement in profits when resources are allocated such that a long-term-oriented, forward-looking metric, such as CLV, is maximized (Venkatesan and Kumar 2004). In contrast, a regression-based estimate projects historic marketing costs into the future while accommodating for any general trends that may exist over time. We conduct an empirical comparison of the performance of the various cost allocation rules for computing forward-looking metrics in selecting profitable customers.

Therefore, the objectives of our research study are as follows:

- To provide a customer selection framework that accommodates the uncertainty inherent in predicting customer behavior,
- To compare the proposed customer selection framework with the collaborating firm's current customer selection framework, and
- To evaluate the total profit implications of the various customer-level cost allocation rules.

The collaborating firm in this study is a large multinational firm that sells high-technology products and services. The customer database of the organization focuses on business-to-business (B2B) customers. In the next section, we provide a background on the customer selection process in the collaborating firm and review the relevant literature on customer selection. In the subsequent section, we illustrate the CLV formulation and the proposed customer selection process. We then discuss the model framework for measuring CLV, the data used to estimate CLV, and the results from model estimation. Following this, we discuss the results from an empirical comparison of the various customer selection procedures. Finally, we provide guidelines for implementation, derive implications based on the results, list the limitations of our study, and identify venues for further research.

BACKGROUND

Customer Selection in the Collaborating Firm

Each year, the collaborating firm proactively contacts its customers through multiple channels, such as through a salesperson, by direct mail (including promotional catalogs and e-mail), and by telephone.[1] As in several other major firms, the marketing department is assigned a budget that can be used to contact current customers. The marketing budget is determined each year on the basis of the performance of the marketing department the previous year, in addition to several other factors. Faced with a limited annual marketing budget, the firm can proactively contact only a fraction of its current customers and therefore must select customers to target with marketing communication every year. Various marketing factors, including product/service innovation, product/service pricing, mass-market advertising, and individual customer contacts, are expected to affect customer profitability. Among these factors, the level of customer contacts has the highest scope for customization

[1]Proactive contacts refer to the firm initiating a contact with a customer without the customer requesting any information.

across customers and is the focus of differential resource allocation for managing customer profitability at the collaborating firm. The firm primarily designs its customer selection strategy as follows:

- The customers are scored on their expected spending potential (ESP) in the following year.
- The customers are then placed into segments according to their scores. The segments are classified as "high potential," "medium potential," and "lowest potential."
- The size of the segments is determined on the basis of the current budget constraints for marketing communication.
- With regard to proactive contacts, customers in the high-potential segment are prioritized over customers in the medium-potential segment, and so forth, until all the resources allocated for the year are used.

ESP

The collaborating firm computes ESP from a Heckman two-stage model (Krishnamurthi and Raj 1988). In this model, a logistic regression is used to predict whether a customer will purchase from the firm in period t. Given purchase in period t, a linear regression (adjusted for the selectivity bias) is used to predict the revenue the firm expects from the customer in period t. The various drivers used to predict ESP can be classified into customer purchase behavior, customer characteristics, and marketing communications. As part of this study, the collaborating firm would use its currently used procedure to provide us with the ESP score for each customer. However, under the current procedure, the firm computes only aggregate-level coefficients. The customer selection process is based on each customer's ESP in the next year. When the customers are selected, they are contacted through the multiple channels until they make a purchase. The level of contacts in each channel is inversely proportional to the unit cost of communication in each channel. Specifically, the level of contacts is highest through direct mail, followed by telephone and then salesperson.

The firm recognizes that there is a large variation across customers, even within the highest-potential segment, in the number of contacts through salespeople, telephone sales, and direct mail required for obtaining a response. The firm intends to design marketing communication strategies that recognize CLV, and a key objective of the firm is to improve its forward-looking perspective in managing customers through optimization of its marketing resource allocation strategy for each customer.

Extant Literature on Customer Selection

Previous research has indicated that customers selected on the basis of a profit maximization approach yield the highest response rates to direct mail catalogs relative to common selection techniques, including CHAID (Bult and Wansbeek 1995), and explicitly accounting for the higher level of aggregation in zip code–based variables leads to improved accuracy in targeting prospects (Steenburg, Ainslie, and Engerbretson 2003). The objective of these studies is to determine the optimal number of customers to select to maximize the total expected profits. The level of marketing resources required for the direct mail campaign is determined by the number of customers selected rather than externally, as is the case for the collaborating firm in our study. Although these previous studies model whether a customer/prospect would respond to a marketing communication, they do not consider the heterogeneity in the revenues a customer provides or the heterogeneity in marketing costs.

Another set of studies that explicitly account for heterogeneity in customer revenues when selecting customers finds that customers selected on the basis of CLV provide more profits than customers selected on the basis of other metrics (Reinartz and Kumar 2003; Venkatesan and Kumar 2004). These studies compare the profits obtained from the top 15% of the customers selected on the basis of a customer metric, but they do not account for a budget constraint or heterogeneity in marketing costs. In this study, we intend to fill these voids in the literature and also address the issues faced by the collaborating firm.

CLV FORMULATION

We measure CLV using the "always-a-share" approach because it is more appropriate for the noncontractual setting of our collaborating firm (Rust, Lemon, and Zeithmal 2004; Venkatesan and Kumar 2004). Thus, in this approach, we measure CLV by predicting customers' purchase patterns over a reasonable period, not when customers will terminate their relationship with the firm. Given predictions of quantity purchased, purchase timing, and variable costs, the CLV formulation we use in this study can be represented as follows:

$$(1) \qquad CLV_i = \sum_{j = T^*+1}^{T^*+T_i} \frac{\hat{Q}_{i,j} \times M}{(1+r)^{\hat{t}_{i,j}}} - \sum_{t=1}^{n} \frac{\sum_q c_{i,q,t} \times x_{i,q,t}}{(1+r)^{t-1}},$$

where

CLV_i = lifetime value of customer i,
$\hat{Q}_{i,j}$ = the predicted purchase quantity for customer i in purchase occasion j,
M = the contribution margin or gross profits for a single item,
r = the discount rate for money (set at 1.25% monthly rate in our study),
$c_{i,q,t}$ = the unit marketing cost for customer i in channel q in year t,
$x_{i,q,t}$ = the number of contacts to customer i in channel q in year t,
$\hat{t}_{i,j}$ = the predicted period of purchase for customer i for the jth purchase occasion,
n = the number of years to forecast (three in this case),
T* = the current time period, and
T_i = the predicted number of purchases made by customer i until the end of three years after T*.

When computing CLV, we assume that there is a yearly allocation of resources (as is the case in the collaborating firm) and that the cost allocation occurs at the beginning of the year (the present period). Thus, the cost allocation in the first year need not be discounted, the cost allocation in the second year needs to be discounted for one year, and so on. Thus, we raise the denominator in the cost function calculation to current year – 1 (i.e., t – 1).

BAYESIAN DECISION THEORY–BASED CUSTOMER SELECTION

Bayesian decision theory is ideally suited for application to problems in which a decision must be made with substantial parameter or modeling uncertainty. Bayesian decision theory postulates that there are actions (e.g., setting marketing decision variables, selecting customers) a firm can take, that there are uncertain states (e.g., quantity and timing of purchases by customers), and that the combination of actions and states results in consequences (e.g., profits). Bayes' theorem combines prior distributions for the states with data to obtain posterior distributions to reduce the uncertainty about the states. The optimal action maximizes expected profit. Expected profits are computed with respect to the predictive distribution of future states (i.e., quantity and timing).[2]

In our context, the manager faces the decision of determining the level of resources to allocate in each communication channel to maximize future profits. This decision must be made using predictions about future customer behavior, such as purchase timing and quantity. Here, we illustrate the Bayesian decision theory approach for customer selection using CLV. Subsequently, we describe how the suggested customer selection approach is easily amenable for scenarios in which firms decide to use other metrics for customer selection. Our proposed customer selection procedure consists of five steps.

Step 1: Model Specification

A probability model is specified that explains how various aspects of customer behavior, such as purchase timing (t_{ij}) and quantity (Q_{ij}), are driven by marketing decision variables ($x_{i,d}$), covariates ($x_{i,cov}$), and customer-level response parameters (β_i).We represent customer behavior by y_i, which includes both t_{ij} and Q_{ij}. The marketing decision variables include the number of contacts through the salesperson, direct mail, and telephone sales channels. The customer behavior factors the probability model is intended to predict correspond with the CLV formulation (Equation 1).

Step 2: Model Estimation

We use the calibration data to estimate the probability distribution of the unknown response parameters for customer i given the observed customer behavior, the marketing decision variables, and the covariates [$p(\beta_i|y_i, x_d, x_{cov})$]. This is the posterior distribution of the customer-level response parameters and is estimated using Markov chain Monte Carlo (MCMC) methods. The MCMC estimation provides a random sample (sample size = P) of response parameters, $\hat{\beta}_i = \{\hat{\beta}_{i1}, \ldots, \hat{\beta}_{iP}\}$, for each customer that represents the posterior distribution of the response parameters for the customer. This step is based on the Bayesian estimation procedure, in which the Bayes' theorem is used to estimate the posterior distribution of the unknown parameters on the basis of prior assumptions of the unknown parameters and the data (for further details, see Rossi and Allenby 2003).

[2]We thank the guest associate editor for providing this succinct explanation.

Step 3: Predicting CLV

Given the posterior distribution of the response parameters estimated in Step 2 and the predictive distribution of customer behavior, $\hat{y}_i$, we calculate CLV in future periods (CLV_i). The posterior expected value of CLV given a certain level of the decision variable is calculated as the average of the CLV values obtained from each sampled value of the posterior distribution of the model parameters. Specifically, the posterior expected CLV for customer i is calculated as the Monte Carlo average:

$$E[CLV_i(\hat{\beta}_i, x_{i,d}, x_{i,cov})] = \Sigma_{p=1}^{P} CLV_i(\hat{\beta}_i, p, x_{i,d}, x_{i,cov})/P.$$

Step 4: Computing Optimal Marketing Costs and Expected Value of CLV

For each customer i, a genetic algorithm is used to obtain the optimal level of the marketing decision variables ($X^*_{i,d}$) and the optimal marketing cost (MC_i^*) that would maximize the expected value of CLV [$E^*(\hat{m}_{i,1})$]. We compute the optimal marketing cost (MC_i^*), corresponding to the optimal level of marketing decision variables for the ith customer, as follows:

$$MC_i^* = \sum_{q=1}^{Q} \tilde{c}_q \times x_{qi}^*,$$

where $\tilde{c}_q$ represents the cost of making a single contact through channel q and x^*_{qi} represents the optimal level of contacts through channel q for customer i based on the posterior distribution of the response parameters.

Step 5: Customer Selection

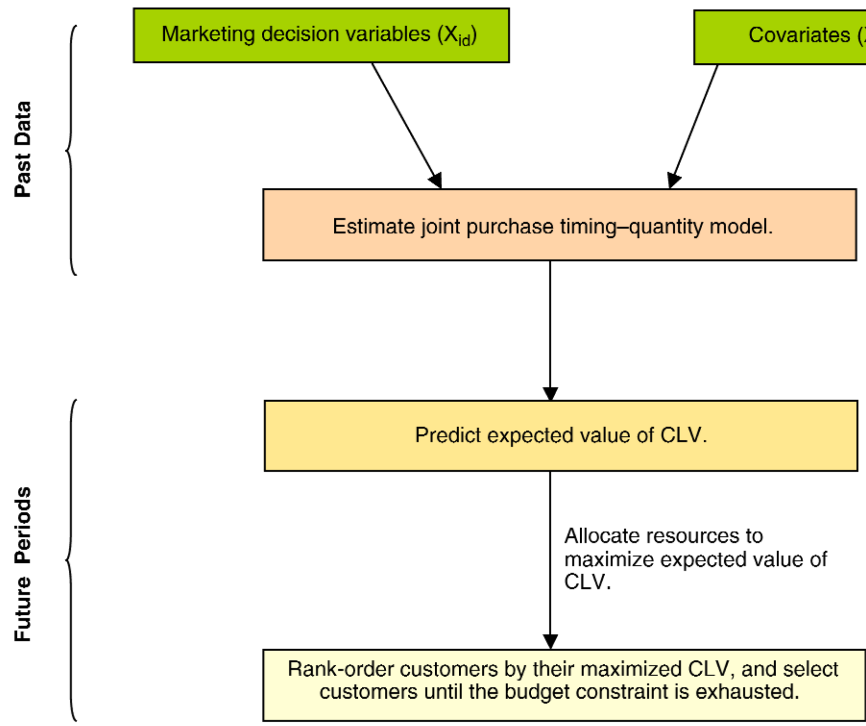
We rank-order customers in descending order of the maximized expected (average) value of their CLV in Step 4. Beginning from the top, we assign customers to the selection set (S*) until the sum of the optimal marketing cost ($\Sigma_i MC_i^*$) for the selected customers equals the budget constraint (BC). (Figure 1 provides a graphical illustration of the proposed customer selection framework.)

MODEL DEVELOPMENT

Determinants of Purchase Timing and Quantity

The commitment–trust theory of relationship marketing (Morgan and Hunt 1994) identifies customer- and supplier-specific factors that are antecedents of future customer activity. Previous research on customer equity and CLV (Bowman and Narayandas 2001; Reinartz and Kumar 2003; Rust, Zeithaml, and Lemon 2004; Venkatesan and Kumar 2004) has provided empirical validation for the antecedents identified in the commitment–trust theory and has found that these antecedents influence a customer's future purchase timing and purchase quantity. Therefore, in line with the commitment–trust theory, we include upgrading, cross-buying, bidirectional communication, returns, and frequency of Web contacts as customer-specific factors that influence a customer's purchase timing and quantity. These customer-specific factors represent the switching costs (upgrading and cross-buying), conflicts (returns), and communication aspects (bidirectional communication and frequency of Web contacts) of a customer–firm relationship.

Figure 1
A CONCEPTUAL FRAMEWORK FOR CUSTOMER SELECTION USING CLV



The supplier-specific factors (or marketing decision variables) include frequency of rich modes of contact and frequency of standardized modes of contact. Similar to previous research (Reinartz, Thomas, and Kumar 2005; Venkatesan and Kumar 2004), we categorize rich modes as contacts through sales personnel and standard modes as contacts through either telephone or direct mail. The collaborating firm also validates that the frequency of marketing contacts is the primary decision variable that firms can customize for each customer every year. Other common marketing decision variables, such as price and promotional discounts, are available to all the customers and do not vary much within a year. On the basis of evidence from previous research for an inverted U-shaped relationship between marketing contacts and customer behavior (Reinartz, Thomas, and Kumar 2005; Venkatesan and Kumar 2004), we include both the linear and the quadratic terms of the marketing decision variables in our model.

Finally, observed customer heterogeneity factors (or control variables) have been found to influence purchase quantity and timing significantly (Allenby, Leone, and Jen 1999; Reinartz and Kumar 2003). Lagged interpurchase time and lagged quantity purchased are used as covariates for both purchase timing and purchase quantity. Both prior interpurchase times and prior quantities are necessary to capture previous customer characteristics. In general, the best customers are expected to purchase larger quantities at shorter intervals, and the worst customers are expected to purchase smaller quantities at longer intervals. For example, prior purchase quantities and prior interpurchase times enable us to evaluate whether customers who make frequent (i.e., smaller prior interpurchase times) large purchases show higher growth in quantity purchased than customers who purchase large quantities but at a lower frequency. We use firm size and industry category as control variables when predicting purchase quantity. For simplicity, we categorize the customer-specific factors and control variables as covariates ($x_{i,cov}$). Table 1 lists the marketing decision variables and covariates used in the study along with the operationalization, expected effect, and rationale for using the variables.

Joint Model for Purchase Timing and Quantity

Recent research (Boatwright, Borle, and Kadane 2003) in a business-to-consumer (B2C) setting (for online grocery retailers) has shown that purchase timing and purchase quantity are dependent on each other. Boatwright, Borle, and Kadane's (2003) joint model, which uses the Conway–Maxwell–Poisson distribution, is not directly applicable to our study. This is because, in general, frequently purchased consumer goods have short interpurchase times (e.g., weeks) and are not categorized by large abrupt changes in the interpurchase times for a customer. In addition, a Poisson distribution assumes a constant, memoryless hazard rate for interpurchase times (i.e., the probability of a customer making the next purchase is highest immediately after the current purchase). Although such an assumption may be appropriate in some B2C settings (e.g., grocery purchases),

Table 1

OPERATIONALIZATION OF MARKETING DECISION VARIABLES, COVARIATES, AND CONTROL VARIABLES

Variable	*Operationalization*	*Expected Effect*	*Rationale*
	Marketing Decision Variables		
Frequency of rich modes of communication	Number of contacts made to the customer by the supplier firm in a month through sales personnel between two observed purchases.	U	Timely communication between parties reduces the propensity of a customer to quit a relationship (Mohr and Nevin 1990; Morgan and Hunt 1994), but too much communication can be detrimental to the relationship (Fournier, Dobscha, and Mick 1997); thus, there is an optimal communication level.
Frequency of standard modes of communication	Number of contacts made by the supplier firm to the customer in a month through telephone or direct mail between two observed purchases.	U	
	Covariates		
Upgrading	Number of upgrades in product purchases until an observed purchase.	–	Customers who upgrade have higher switching costs with each upgrade, which can lead to lower propensity to leave and higher recurrent needs (Bolton, Lemon, and Verhoef 2004).
Cross-buying	Number of different product categories a customer has purchased.	–	Customers who purchase across several product categories have higher switching costs and recurrent needs (Bowman and Narayandas 2001; Reinartz and Kumar 2003).
Bidirectional communication	Ratio of number of customer-initiated contacts to total number of contacts made to the customer (both customer initiated and supplier initiated) between two observed purchases.	–	Two-way communication between parties strengthens the relationship and ensures that the focal firm is recalled when a need arises (Morgan and Hunt 1994).
Returns	Total number of times a return is registered for a customer until the current purchase occasion.	U	Returns provide an opportunity for firms to satisfy customers and ensure repeat purchases (Reinartz and Kumar 2003), but too many purchases can be detrimental to the relationship and can indicate that the firm has not used the return opportunity appropriately.
Frequency of Web contacts	Number of times the customer contacts the supplier through the Internet in a month between two observed purchases.	–	Customers who use online communication want transactions efficiencies, and customers who want to create efficiencies are highly relational and have recurring needs (Grewal, Corner, and Mehta 2001).
Size of the establishment	Number of employees in the customer firm.	+	Control variables that accommodate customer heterogeneity (Niraj, Gupta, and Narasimhan 2001).
Industry category	Standard Industrial Classification category to which the customer firm belongs.	+	

Notes: U = U-shaped curve.

it is too simplistic for the complex decision process characteristic of B2B settings and such B2C settings as apparel retailing. Another commonly used approach for modeling purchase quantity and timing simultaneously is to model the incidence of purchase at each time interval (e.g., weeks) and the purchase quantity given incidence (e.g., Chintagunta 1993). This approach was developed for the modeling of customer behavior for grocery purchases, in which customers make regular and frequent visits to a grocery store. However, such a framework would not be parsimonious in our context, because an average customer would make about four purchases a year; therefore, the number of no-purchase occasions and the sample size for the model framework would be considerably high. Under this scenario, the concomitant mixture framework (Allenby, Leone, and Jen 1999; Venkatesan and Kumar 2004) is more appropriate to model interpurchase times than is the Conway–Maxwell–Poisson distribution (Boatwright, Borle, and Kadane 2003).

We assume that the customers fall into K segments and that purchase timing and quantity are independent given membership in segment k. For example, we assume that there are two segments (K = 2) of customers—heavy (k = 1) and light (k = 2). Customers that belong to Segment 1 are expected to have shorter interpurchase times and to purchase larger quantities than customers in Segment 2. However, given that a customer belongs to Segment 1, his or her current interpurchase time does not determine the current purchase quantity, or vice versa. Modeling segment membership provides an early warning indication of whether a customer is about to enter into a light segment and also accounts for the large abrupt changes in interpurchase times and quantity for a customer. The identification of segments in the proposed model framework is also an attractive factor for CRM applications because the primary objective of CRM is to customize marketing actions to each individual or to each customer segment.

We model customers' interpurchase times using the generalized gamma distribution.[3] The generalized gamma model also accommodates the commonly used exponential distribution for interpurhcase times (Reinartz and Kumar 2000, 2003; Schmittlein, Morrison, and Colombo 1987). For a given customer i, we assume that the interpurchase time (measured in months) for the jth purchase occasion is a distributed generalized gamma (GG); that is,

$$t_{ij} \sim GG(\alpha, \lambda_{ij}, \gamma) = \frac{\gamma}{\Gamma(\alpha)\lambda_{ij}^{\alpha\gamma}} t_{ij}^{\alpha\gamma - 1} e^{-(t_{ij}/\lambda_{ji})^{\gamma}}. \tag{2}$$

The parameters α and γ establish the shape of the distribution (see McDonald and Butler 1990), and λ_{ij} is the purchase rate parameter or the scale of the interpurchase time distribution. Similar to Allenby, Leone, and Jen's (1999) and Boatwright, Borle, and Kadane's (2003) approaches, we model the purchase rate parameter as follows:

$$\lambda_{ij} = \lambda_i \eta_1^{t^*_{ij-1}} \eta_2^{t^*_{ij-2}} \eta_3^{Q^*_{ij-1}} \eta_4^{Q^*_{ij-2}}, \tag{3}$$

where

$\eta_1 - \eta_4$ = the coefficients for the covariates,
Q^*_{ij-1} = the lagged log of purchase quantity for customer i,
Q^*_{ij-2} = the two-period lagged log of purchase quantity for customer i,
t^*_{ij-1} = the lagged log of interpurchase time for customer i, and
t^*_{ij-2} = the two-period lagged log of interpurchase time for customer i.

Under this formulation, customer i's purchase rate at each purchase occasion j is a function of the customer's individual-specific purchase rate parameter (λ_i), the previous interpurchase times (t_{ij-1}, t_{ij-2}), and the previous quantities purchased (Q_{ij-1}, Q_{ij-2}). The term λ_i captures cross-customer variation, and the covariates, lagged interpurchase times, and lagged quantity purchased capture temporal variation within a customer. The covariates capture the effect of the level of previous purchases (lagged purchase quantities) and the frequency of previous purchases (lagged interpurchase times) on the timing of the current purchase occasion. These covariates can also capture the trend and cycles in consumption patterns over time. The parameters (η) are restricted to be greater than zero, and they measure the impact of the covariates (i.e., previous interpurchase times and previous quantities purchased) on the current purchase rate parameter. In this case, $\eta > 1$ implies a positive effect of a covariate on the purchase rate parameter (i.e., reduced interpurchase time), $\eta = 1$ implies no effect, and $0 < \eta < 1$ implies a negative effect. For example, if η_1 is less than 1, then the longer the interpurchase time in the previous purchase occasion, the smaller is the purchase rate parameter, and the longer is the current interpurchase time. If the covariates are measured in logarithms, the coefficients (ηs) can be interpreted as the percentage change in the expected interpurchase time for a 1% change in the covariate. We use the conjugate inverse generalized gamma distribution for the individual-specific purchase rate parameter λ_i,

$$\lambda_i \sim IGG(\nu, \theta, \gamma) = \frac{\gamma}{\Gamma(\nu)\theta^{\nu\gamma}} \lambda_i^{-\nu\gamma - 1} e^{-(1/\theta\lambda_i)^{\gamma}}. \tag{4}$$

For modeling purchase quantity, we need to address endogeneity issues that arise when prior quantity purchased is included as an independent variable to predict current purchase quantity. In panel data models with lagged dependent variables, the endogeneity in formulation can be alleviated with a one-period difference in the dependent variable and a two-period lagged dependent variable as an independent variable (Baltagi 1998). We use the growth in quantity from purchase occasion j – 1 to j as the dependent variable and the quantity in purchase occasion j – 2 as an independent variable. Similar to the formulation for the purchase rate parameter (λ_{ij}), we include logs of previous interpurchase times and previous quantities as covariates in the quantity model. These variables capture the effect of previous buying behavior on growth in quantity purchased. We include the two- and three-period lagged purchase quantities and the one- and two-period lagged interpurchase times as independent variables. In line with beliefs about the collaborating firm and evidence from extant literature (Reinartz and Kumar 2003; Venkatesan and Kumar 2004), we also include firm size and industry category as independent variables to capture heterogeneity in purchase quantities across customers. The purchase quantity for a customer i is provided by[4]

$$\Delta Q_{i,j} = \delta_{i,0} + \delta_1 \times Q^*_{i,j-2} + \delta_2 \times Q^*_{i,j-3} + \delta_3 \times t^*_{i,j-1} + \delta_4 \times t^*_{i,j-2} + \delta_5 \times \text{size}_i + \sum_{c=1}^{C} \delta_{6,c} \times \text{Ind}_{ci} + e_{ij}, \tag{5}$$

where

$\Delta Q_{i,j}$ = the growth in purchase quantity between purchase occasions j – 1 and j,
$\delta_{i,0}$ = a random individual-specific intercept term,
δ^* = δ_1–δ_6 are the coefficients for the covariates,
Q^*_{ij-3} = the three-period lagged log of purchase quantity for customer i,
size_i = the number of employees in the customer firm,
Ind_i = the indicator variable for the industry category of the customer firm,
C = the number of industry categories – 1, and
e_{ij} = the normally distributed random error term with mean zero and variance σ^2.

For both the purchase rate parameter (Equation 3) and the growth in purchase quantity (Equation 5), additional

[3]The generalized gamma distribution provides a better in-sample fit for our data than the exponential, Weibull, gamma, and log-logistic distributions. The results are available on request.

[4]We drop the first purchase occasion of a customer in our calibration data and therefore alleviate the need to make assumptions about Q at time 0.

lagged effects of purchase quantity and purchase timing did not have a significant influence, so we do not include them in our model formulation.

In our model framework, each observation in the data has a distinct probability of belonging to the various subgroups. We determine the probability that a purchase occasion j from a customer i belongs to subgroup k, φ_{ijk}, from a cumulative normal distribution (probit, Φ), which is a function of the marketing decision variables ($x_{i,d}$) and the covariates ($x_{i,cov}$) listed in Table 1. The segment membership probabilities for the case when there are two segments (k = 2) are given by

$$\phi_{i,j,1} = 1 - \Phi(x'_{i,j}\beta_i) \text{ and} \tag{6}$$

$$\phi_{i,j,2} = \Phi(x'_{i,j}\beta_i). \tag{7}$$

Similar to Allenby, Leone, and Jen (1999), we ensure identification by imposing the restriction that $\theta_k > \theta_{k-1} \ldots > \theta_1$ (the scale parameter of the inverse generalized gamma distribution of λ_I; see Equation 4). We allow the parameters β_i to be heterogeneously distributed across customers by a random effects specification

$$\beta_i \sim \text{normal}(\bar{\beta}, \Sigma_\beta). \tag{8}$$

We assume that the interpurchase time and quantity for the jth observation for customer i are independent, given segment membership (k). Therefore, for each customer, the expected time until the next purchase is a weighted sum (the weights are stochastically determined by the probit function of the covariates) of their predictions of expected time to next purchase from each subgroup and is given by

$$t_{ij} \sim \sum_k \phi_{ijk} GG(\alpha_k, \lambda_{ijk}, \gamma_k). \tag{9}$$

Similarly, for each customer, the expected quantity is given by

$$\Delta Q_{i,j} = \sum_k \phi_{ijk}(\delta_{i0k} + \delta_{1k} \times Q^*_{ij-2} + \delta_{2k} \times Q^*_{ij-3} + \delta_{3k} \times t^*_{ij-1} + \delta_{4k} \times t^*_{ij-2} + \delta_{5k} \times \text{size}_i + \sum_{c=1}^{C} \delta_{6ck} \times \text{Ind}_{ci} + e_{ijk}). \tag{10}$$

Finally, the likelihood function (for the joint purchase timing–quantity model) is specified as

$$L = \prod_{i=1}^{n}\prod_{j=1}^{J_i}\sum_{k=1}^{K} \phi_{ijk}\Big[f_k\Big(t_{ij}|\alpha_k, \lambda_{ijk}, \gamma_k\Big) \times p(\Delta Q_{ij}|\delta_{i,k}, \delta^*_k, \sigma^2_k)\Big]^{c_{ij}} S_k(t_{ij}|\alpha_k, \lambda_{ik}, \gamma_k)^{(1-c_{ij})}, \tag{11}$$

where

$f(t_{ij}|\alpha, \lambda_i, \gamma)$ = the density function for the generalized gamma distribution (in other words, the probability of the jth purchase for customer i occurring at period t, given $\alpha, \lambda_i, \gamma$);

$S(t_{ij}|\alpha, \lambda_i, \gamma)$ = the survival function for the generalized gamma distribution (in other words, the probability of the jth purchase for customer i occurring at a period is greater than t, given that the jth purchase has not occurred until time t, given $\alpha, \lambda_i, \gamma$);

$p(\Delta Q|\delta_i, \delta^*, \sigma^2)$ = the density function for purchase quantity (Equation 9); and

c_{ij} = the censoring indicator, where c_{ij} = 1 if the jth interpurchase time for the ith customer is not right censored and c_{ij} = 0 if the jth interpurchase time for the ith customer is right censored.

Given the model parameters, the predictive distribution from Equations 9 and 10 provides the interpurchase times and purchase quantity in the forecast periods. We can then use these predictions to compute the customer metrics. Web Appendix WA (see http://www.marketingpower.com/content84060.php) provides the MCMC estimation algorithm for the proposed joint model, the priors used in the estimation, and the algorithm used to predict interpurchase time and purchase quantity in the forecast period. We also compare our proposed framework for jointly modeling purchase timing and quantity with Boatwright, Borle, and Kadane's (2003) framework as a benchmark. Web Appendix WB (see http://www.marketingpower.com/content84060.php) describes how we adapt Boatwright, Borle, and Kadane's framework to our context and compares the performance of our model with the benchmark model.

DATA

The collaborating firm in this study sells several high-technology products and services to business customers. The firm's products typically require maintenance and frequent upgrades; these provide the variance required to model purchase timing and purchase quantity. For our analyses, we use two cohorts of customers, Cohort 1 and Cohort 2. Customers were assigned to Cohort 1 (Cohort 2) if their first purchase with the manufacturer was made in the first quarter of 1997 (first quarter of 1998). For Cohort 1 (Cohort 2), we use the first 48 months (36 months) of data as the calibration sample to estimate the joint purchase timing–quantity model. We use the next 36 months of data for both Cohort 1 and Cohort 2 as the holdout sample to evaluate the effectiveness of the customer selection strategies. In our samples, we removed customers who had missing values for either rich or standardized modes of communication. We also restricted our sample to customers who made at least five purchases. Overall, we removed 20% of the original cohort of customers for our analyses. We also removed the first three purchases for each customer to avoid missing values for the lagged covariates. Then, we randomly sampled 238 customers from Cohort 1 and 210 customers from Cohort 2, which resulted in 4326 and 3521 purchase occasions for Cohorts 1 and 2, respectively. We used each purchase occasion as an observation (j) for the joint purchase timing–quantity model (Equation 11).

We measure the marketing decision variables (x_d) as the number of contacts made through the rich and standardized modes in a month between two observed purchases (j and j – 1) for the joint purchase timing–quantity model. We recognize that all the contact information for each customer may not be recorded in the data. However, we expect that the extent of missing information does not vary systematically from one customer to another. Because our focus is on relative profits across customers and across metrics, the missing information should not substantially influence our conclusions.

The covariates (x_{cov}) (listed in Table 1) can be classified as cumulative and current effects. The cumulative effects covariates include cross-buying, upgrading, and returns. Their values represent the total number of different products (for cross-buying) or upgrades the customer has purchased or the returns the customer has made since first purchase until the current observation. As we described previously, the current observation represents a purchase occasion in the joint purchase timing–quantity model. The current effects covariates include bidirectional communication and frequency of Web contacts. We calculate the current effects covariates on the basis of customer activities between the previous purchase occasion (j – 1) and the current purchase occasion (j). The covariates firm size and industry category (used to predict purchase quantity) vary across customers but do not change across purchase occasions for a customer.

All the marketing decision variables and covariates for the segment membership functions (Equations 6 and 7) used in our analyses are lagged variables to address the possibility of endogeneity. Specifically, for observed purchase j, the cumulative effects antecedents represent activity of the customer since first purchase until observed purchase j – 1. Similarly, for observed purchase j, the current effects antecedents and covariates represent customer (or supplier) activity between observed purchase j – 2 and j – 1. In addition to addressing the potential endogeneity issues, we address the lagged variables because when selecting customers for targeting, the manager would need to predict the performance of a customer on each metric based on his or her knowledge about the customer's activities until the current time period. Therefore, a model that is built using lagged values of the marketing decision variables and the covariates would allow the manager to use the estimated coefficients to predict future values of CLV.

RESULTS FROM MODEL ESTIMATION

Model Comparison

We use the aggregate log conditional predictive ordinate (CPO) for evaluating the in-sample fit of the models (Gelfand and Dey 1994). Similar to the log-likelihood, a higher value of the aggregate log CPO is interpreted as a better model fit. To evaluate the predictive accuracy of the models, we reestimate the model using all but the last observation in the calibration sample for each customer. We then use the model estimates to predict the last observation—both purchase timing and quantity. The mean absolute deviation (MAD) between the predicted and the observed values provides an estimate of the predictive accuracy of the model. We also compare the predictive accuracy of each model with the accuracy of a naive estimate, which is defined as the average of the interpurchase time and purchase quantity in the calibration sample (excluding the last observation). We then compute the relative absolute error (RAE) of a model as the ratio of the MAD of the model to the MAD of the naive estimate (Armstrong, Morvitz, and Kumar 2000). We provide the results of the comparison of in-sample fit and the predictive accuracy over the different models in Table 2.

We compare the joint purchase timing–quantity model illustrated here (the proposed model) with Venkatesan and Kumar's (2004) model framework (the base model) and Boatwright, Borle, and Kadane's (2003) framework (the benchmark model). We do not check other, simpler model frameworks because Venkatesan and Kumar (2004) find that their model framework (the base model) provides a better in-sample fit and predictive accuracy than other models, such as (1) a finite mixture model framework; (2) a generalized gamma model with no segments, no marketing decision variables, and no covariates; and (3) a generalized gamma model with segments but no marketing decision variables or covariates. We obtained similar results for both Cohort 1 and Cohort 2; thus, we report the results from Cohort 1 for simplicity. For the base and the proposed models, we used 50,000 iterations of the MCMC algorithm (see Web Appendix WA at http://www.marketingpower.com/content84060.php).[5] We used the initial 30,000 iterations as burn-in and the last 20,000 iterations as the posterior sample. The autocorrelation function revealed that every 10th sample in the posterior distribution is unrelated, and there-

[5]Estimation details for the benchmark model appear in Web Appendix WB (see http://www.marketingpower.com/content84060.php).

Table 2

MODEL EVALUATION

A: Determination of Number of Segments for CLV Model[a]		
Number of Segments	*Base Model*	*Proposed Model*
1	–6859	–6032
2	–6792	–5945
3	–6801	–6012
4	–6848	–6143

B: Comparison of Model Performance	
Aggregate Log CPO	
Base model	–6792
Proposed model	–5945
Benchmark model	–6542
MAD[b]	
Base model	Purchase time = 2.8
	Purchase quantity = 7.9
Proposed model	Purchase time = 1.8
	Purchase quantity = 6.1
Benchmark model	Purchase time = 3.0
	Purchase quantity = 7.3
RAE	
Base model	Purchase time = .61
	Purchase quantity = .55
Proposed model	Purchase time = .40
	Purchase quantity = .42
Benchmark model	Purchase time = .66
	Purchase quantity = .51

[a]Reported values are the aggregate log CPO.
[b]Purchase time is measured in months.

fore we used 2000 samples (every 10th from the 20,000 available in the posterior sample) in the posterior distribution to make our inferences.

We estimated four versions of both the base and the proposed models by varying the number of segments (k). We compared the aggregate log CPO of the four versions, and we provide the results in Table 2, Panel A. Table 2, Panel A, shows that the aggregate log CPO is the highest for the model with two segments for both the base and the proposed models, and therefore we use the model version with two segments for further analyses. In our study, we observe that the aggregate log CPO values indicate that accommodating the dependence between purchase timing and quantity leads to better in-sample fit to the data (aggregate log CPO for proposed model = –5945) than a model that assumes that purchase timing and quantity are independent (aggregate log CPO for base model = –6792). The better in-sample fit also translates into better predictive accuracy. The MAD for predicting purchase time is equal to 2.8 months for the base model and 1.8 months for the proposed model. Similarly, the MAD for predicting purchase quantity is equal to approximately 7.9 months for the base model and 6.1 months for the proposed model.

For the benchmark model, MADs for predicting purchase time and quantity are 3.0 and 7.3, respectively (see Table 2, Panel B). For each model, we also compared the predictive accuracy with a naive estimate. The naive estimate was the average interpurchase time and purchase quantity calculated over the period of the calibration sample. The mean interpurchase time and purchase quantity in the calibration data are 4.1 and 23 months, respectively. The RAE for the base model is.61 for purchase time and .55 for quantity purchased. The RAE for the proposed model is .40 for purchase time and .42 for quantity purchased. Finally, the RAE for the benchmark model is .66 and .51 for purchase time and purchase quantity, respectively.

The RAE measures indicate that all the models (the base, the proposed, and the benchmark models) provide better predictive accuracy than simple heuristics (given that the RAEs are less than 1), in addition to providing a framework for linking customer behavior to marketing actions. Overall, the proposed model framework is more appropriate than other alternatives for modeling purchase timing and purchase quantity in a B2B setting for high-technology products.

Influence of Decision Variables and Covariates

We report the results of the proposed model in Table 3. The reported values are the posterior means and standard deviations (within parentheses). A parameter is considered "not significant" if a zero exists within the 2.5th percentile and the 97.5th percentile values of the posterior distribution for that parameter.

For the segment membership model, we report the mean of the random effects distribution, $\bar{\beta}$, in Equation 8. The average interpurchase time and quantity for customers who belong to Segment 1 is 2 and 41 months, respectively, and the average interpurchase time and purchase quantity for customers who belong to Segment 2 is equal to 5 and 14 months, respectively. Therefore, we call Segment 1 the "heavy segment" (lower interpurchase time and larger quantity) and Segment 2 the "light segment" (higher interpurchase time and smaller quantity). The parameters reported in Table 3, Panel A, determine the probability that an observation will belong to the heavy segment (Equation 6). The mass point reported in Table 3, Panel A, indicates that approximately 56% of the observations belong to the heavy segment.

The parameters of the generalized gamma distribution (used to model interpurchase times) and the parameters of the quantity model and the purchase rate parameter (see Table 3, Panel B) for each segment are substantially different from each other. This implies that the model can accommodate differences in interpurchase times and purchase quantities between observations in heavy and light segments, and it also substantiates the choice of a concomitant mixture framework for the joint modeling of interpurchase times and purchase quantity in a B2B setting.

Impact of marketing decision variables. Similar to previous research (Reinartz, Thomas, and Kumar 2005; Venkatesan and Kumar 2004), we also find that marketing decision variables have a nonlinear, inverted U-shaped influence on customer behavior. The linear terms for both rich and standardized modes are positive and significant (frequency of rich modes = 3.5, and frequency of standard modes = 5.3), whereas the quadratic terms for both rich and standard modes are negative and significant (square of frequency of rich modes = –1.9, and square of frequency of standard modes = –1.1). For both rich and standard modes, the parameter estimates indicate that until a certain threshold, an increase in the frequency of touches increases the probability that an observation will belong to the heavy segment. However, beyond the threshold, increasing the frequency of touches decreases the probability that an observation will belong to the heavy segment. Figure 2 provides a distribution of response coefficients across customers for the frequency of rich modes of contact (both the linear and the quadratic terms). We observe a skewed distribution with long tails for the response coefficients of both the frequency of rich modes (the linear term) and the square of frequency of rich modes (the quadratic term), implying that there is potential for increased profits from customizing the level of contacts to each customer. We also observe a similar distribution for the response coefficients for the frequency of standard modes.

Impact of covariates. Similar to the expectations in the customer management literature (Bolton, Lemon, and Verhoef 2004), we find that an increase in upgrading, cross-buying, bidirectional communication, and frequency of Web contacts leads to an increase in the probability that an observation will belong to the heavy segment. We also find that returns have a nonlinear, inverted U-shaped influence on the likelihood that an observation will belong to the heavy segment.

Overall, we find that shorter interpurchase times in the previous purchase occasions and higher purchase quantities in the previous purchase occasions are associated with a higher purchase rate parameter, a shorter interpurchase time, and a higher growth in purchase quantity. All the lagged effects have a significant influence on the purchase quantity and the interpurchase time in the current period for observations that belong to the heavy segment. Only the interpurchase time for the previous purchase occasion and the quantity purchased in the previous two purchase occasions have a significant influence on the purchase rate parameter for observations that belong to the light segment.

TABLE 3
RESULTS FROM MODEL ESTIMATION

A: Segment Membership[a]					
Marketing Decision Variables			*Covariates*		
Frequency of rich modes	3.5	(.32)	Upgrading	2.5	(.50)
(Frequency of rich modes)^2	–1.9	(.15)	Cross-buying	3.1	(.29)
Frequency of standard modes	5.3	(.98)	Bidirectional communication	2	(.49)
(Frequency of standard modes)^2	–1.1	(.98)	Returns	2.5	(.40)
			(Returns)^2	–4.2	(.23)
			Frequency of Web contacts	3.2	(.95)
			Mass point	.56	(.10)

B: Purchase Timing: Generalized Gamma Distribution				
	Heavy Segment		*Light Segment*	
α	9.32	(.54)	6.20	(.11)
ν	1.32	(.12)	2.1	(.07)
θ	45.72	(3.20)	37.82	(.45)
γ	1.30		1	
Lagged log of interpurchase time	.88	(.42)	.93	(.72)
Two-period lagged log of interpurchase time	.95	(.85)	n.s.	
Lagged log of quantity	1.82	(.92)	1.45	(1.12)
Two-period lagged log of quantity	1.75	(.89)	1.35	(1.23)

Purchase Quantity: Panel Regression				
	Heavy Segment		*Light Segment*	
Intercept[b]	–.19	(.01)	.05	.03
Two-period lagged log of quantity	.79	(.10)	.35	.15
Three-period lagged log of quantity	.35	(.08)	n.s.	
Lagged log of interpurchase time	–2.50	(.11)	–2.30	.07
Two-period lagged log of interpurchase time	–2.50	.10	n.s.	
Size	.15	(.01)	n.s.	
Aerospace	.26	(.09)	.13	(.02)
Financial services	n.s.		n.s.	
Manufacturing	n.s.		n.s.	
Technology	.45	(.06)	.12	.04
Consumer packaged goods	n.s.		.42	(.06)
Travel	.26	(.07)	n.s.	
Government	n.s.		n.s.	
σ^2	.32	(.02)	2.10	(.94)

[a]Values reported are posterior means of the mean of the random effects specification, $\bar{\beta}$, and values in parentheses are posterior standard deviations.
[b]Values reported are the means of the individual level intercept term, and the standard deviation across customers is reported in parentheses.
Notes: n.s. = not significant (i.e., a zero exists between the 2.5th percentile and the 97.5th percentile values of the posterior distribution). All variables are significant at least at α = .05 unless specified otherwise.

Regarding growth in purchase quantity, only the two-period lagged purchase quantity and lagged interpurchase time have a significant influence for observations that belong to the light segment. In our sample, among observations that belong to the heavy segment, larger customers (size) and customers that belong to the aerospace, technology, and travel industries have a higher growth in purchase quantity than other customers. Among observations that belong to the light segment, customers that belong to the aerospace, technology, and consumer packaged goods industries have a higher growth in purchase quantity than other customers.

COMPARISON OF SELECTION CAPABILITY

In this section, we compare (1) the selection capability of CLV (using a status quo cost allocation rule) with the capability of the ESP metric used by the collaborating firm and (2) the selection capabilities of the various cost allocation rules (i.e., the status quo, the optimal, and the regression-based cost allocation rule) used for calculating CLV. In general, the selection exercise proceeds through the following three steps:

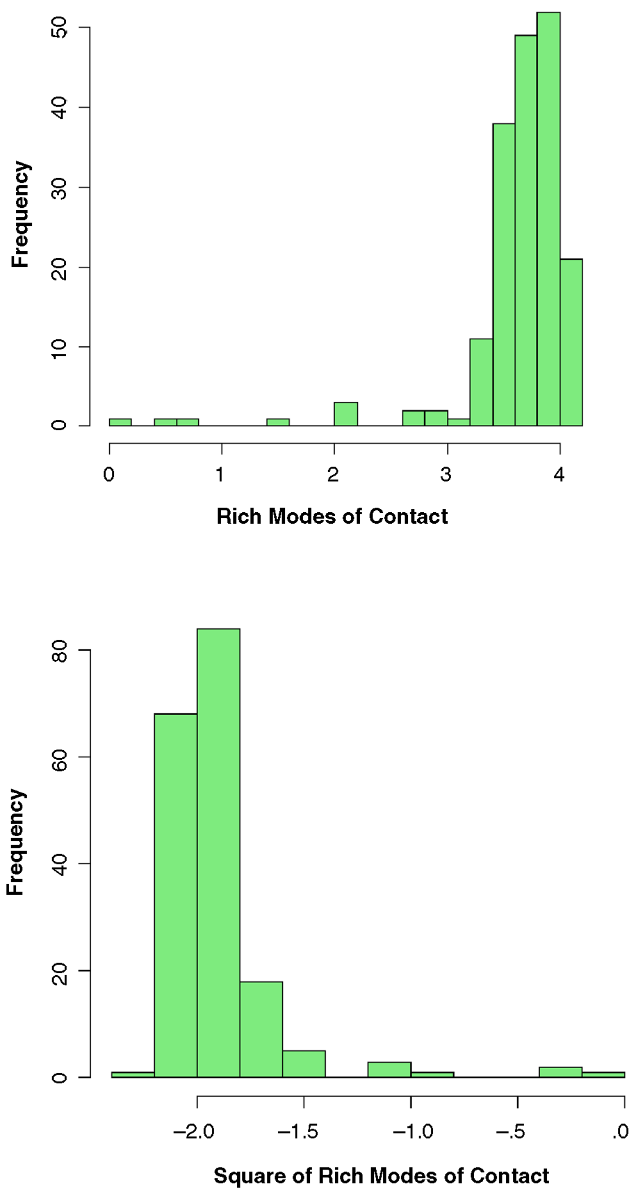
Step 1: We use the estimates obtained from the calibration sample to score (the collaborating firm provided the scores for ESP) and sort the customers on the metric in the holdout period.

Step 2: We pick the top m customers until the estimated total cost of serving the customers exceeds the budget constraint.

Step 3: For selected customers, we obtain the observed net profits (contribution margin less costs) in the holdout period.

The higher the observed net profits in the holdout period, the better is the selection capability. We present the results from using the proposed model here because it provides better in-sample fit and predictive accuracy than the base model and the benchmark model. We set the budget constraint at 70% of the average of the total annual cost of serving all the customers in the calibration sample over the most

Figure 2
DISTRIBUTION OF RESPONSE COEFFICIENTS FOR MARKETING DECISION VARIABLES



recent three years (1999–2001). Therefore, the total budget available for serving all the customers in Cohort 1 is approximately $100,000. When predicting multiple periods ahead (for calculating CLV), we recursively use the predicted values for a particular purchase occasion as an independent variable for predicting the next purchase occasion. For example, to predict the fifth purchase occasion after the current period (T*), we use the predicted purchase time for the fourth and third purchase occasions since T* as independent variables for purchase rate parameter (Equation 3) and growth in purchase quantity (Equation 5). Further details on using the model predictions for calculating CLV appear in Web Appendix WA (see http://www.marketingpower.com/content84060.php).

Determining Optimal Marketing Costs

We used a genetic algorithm to obtain the optimal marketing cost for each customer. Specifically, we varied the frequency of contacts through the rich and standard modes for each customer and then calculated the sum of the expected value of CLV of all the customers in the sample. The objective function of the optimization algorithm was to maximize the sum of customers' expected CLV. The optimization algorithm maximized the sum of the expected CLV by varying 476 parameters (i.e., by varying the frequency of contacts through two modes—rich and standard—for the 238 customers in Cohort 1). On the basis of discussions with the collaborating firm, we set the unit costs of serving a customer through the standard modes and rich modes at $6 and $60, respectively. We set the parameters in the genetic algorithm as follows: population size = 200, probability of crossover = .8, probability of mutation = .25, and convergence criteria = the difference in optimal solution over the last 10,000 iterations is less than .1% (for further details on the genetic algorithm, see Venkatesan, Krishnan, and Kumar 2004).

Obtaining a Regression-Based Estimate of Marketing Costs

To obtain a regression-based estimate, we first used annual costs in the calibration sample to estimate the following equation:

$$\text{(12)} \quad \text{Cost}_{i,t} = \zeta_0 + \zeta_1 \times \text{Cost}_{i,t-1} + \zeta_2 \times \text{Cost}_{i,t-2} + \zeta_3 \times \text{Cost}_{i,t-3} + \zeta_4 \times \text{TQ}_{i,t-1} + e_{it},$$

where

$\text{Cost}_{i,t}$ = total cost of contacting customer i in year t;
$\text{Cost}_{i,t-1}, ..., \text{Cost}_{i,t-3}$ = total cost of contacting customer i in year t – 1, t – 2, and t – 3, respectively;
TQ_{it-1} = total quantity purchase by customer i in year t – 1;
ζs = the coefficients to be estimated; and
e_{it} = the error term.

Although the influence of several substantive decision variables can be assessed in the cost equation, our interest is only to obtain a good estimate of future costs using a regression analysis. Therefore, we include only lagged cost and lagged total quantity purchased as independent variables. The lagged cost variables (cost_{it-1}, cost_{it-2}, cost_{it-3}) capture the trend in the cost of serving a customer over time and also accommodate any factor that may affect the cost of serving a customer. The total quantity purchased in the previous year (TQ_{it-1}) adjusts the trend in cost of serving a customer to the level of the customer's most recent purchases.

The regression of lagged marketing costs on current marketing costs provided an R-square of approximately .75. We then used the following equation as an estimate of the future cost of serving a customer:

$$\text{(13)} \quad \text{Cost}_{i,t+1} = \hat{\varsigma}_0 + \hat{\varsigma}_1 \times \text{Cost}_{i,t} + \hat{\varsigma}_2 \times \text{Cost}_{i,t-1} + \hat{\varsigma}_3 \times \text{Cost}_{i,t-2} + \hat{\varsigma}_4 \times \text{TQ}_{it-1},$$

Table 4
EVALUATION OF CUSTOMER SELECTION CAPABILITY

	CLV			*ESP*
Metric Used to Select Customers	*Status Quo Marketing Costs (Cell A)*	*Optimal Marketing Costs (Cell B)*	*Regression-Based Estimate of Costs (Cell C)*	*Status Quo Marketing Costs (Cell D)*
Number of customers	151	152	152	182
NPV of profits	1901	2425	1994	1641
NPV of costs	318	283	294	400

Notes: NPV of profits is that which we observed in the holdout sample over three years, and NPV of costs is that which we observed in the holdout sample over three years. We use a monthly discount rate of 1.25, and both are in $1,000s.

where $\hat{\varsigma}_s$ are the coefficients estimated in Equation 11.

In Table 4 we report the NPV of profits and costs observed in the holdout sample for the customers selected under each scenario. We use a monthly discount rate of 1.25% for calculating the NPV.

Comparison with the Current Practice in the Collaborating Firm

As we described previously, the collaborating firm uses a customer's ESP to select customers. The collaborating firm provided us with the ESP scores for the customers using the calibration sample, which we then used in our selection exercise. Currently, the firm does not estimate customer-specific response coefficients, uses status quo marketing costs as an estimate of future costs, and does not use a Bayesian decision theory–based approach for customer selection. To control for the influence of the cost allocation rule, we compare the performance of customer selection using (1) CLV and status quo marketing costs (Cell A) and (2) ESP and status quo marketing costs (Cell D).

We observe that fewer customers are selected in Cell A (151) than in Cell D (182). The cost of serving customers selected in Cell A ($318,000) is lower than the cost of serving customers selected in Cell D ($400,000).[6] Finally, the customers selected in Cell A provide higher profits ($1.9 million) than customers selected in Cell D ($1.6 million). Overall, although the CLV metric selects fewer customers than the ESP metric, the customers selected by CLV are more profitable (profit per selected customer is $12,582) than those selected by ESP (profit per selected customer is $8,791). This comparison provides strong support for using CLV for customer selection over the collaborating firm's current practice of using ESP.

Comparison of Cost Allocation Rules

Given the support for using CLV for customer selection, we now compare the capability of the various cost allocation rules when using CLV (i.e., Cells A, B, and C). Among the various cost allocation rules, using CLV with optimal marketing costs (Cell B) corresponds to the Bayesian decision theory–based customer selection framework we propose herein.

Status quo versus regression-based marketing costs. We observe that the status quo cost allocation rule (Cell A) and the regression-based marketing cost (Cell C) select a similar number of customers. The cost of serving customers selected in Cell C ($294,000) is lower and closer to the expected costs (i.e., the budget constraint) than the cost of serving customers in Cell A ($318,000). The customers selected in Cell C also provide higher profits ($1.99 million) than those selected in Cell A ($1.9 million). Overall, our selection exercise provides strong support for using a regression-based estimate (a forward-looking procedure) of marketing costs when using CLV for customer selection.

Optimal marketing costs versus regression-based estimates. Both the proposed Bayesian decision theory–based procedure (i.e., Cell B) and the regression-based cost allocation rule (Cell C) select similar number of customers. The cost of serving customers selected in Cell B is also similar to the cost of serving customers selected in Cell C. However, the customers selected in Cell B provide higher profits ($2.42 million) than those selected in Cell C.

The results indicate that managers need to take a forward-looking perspective for managing customer costs, and they imply that the rank ordering of customers based on the maximized CLV (using optimal marketing costs) provides better targeting of profitable customers than the rank ordering of customers based on a CLV measure obtained using a simple regression-based estimate of future costs. A possible reason for optimal marketing costs providing better targeting is that the optimal marketing costs enable managers to dynamically update their estimates of future customer costs according to customer responsiveness to historic marketing communication, whereas the regression-based estimate simply projects historic marketing costs into the future while accommodating any general trends that may exist over time. However, further research that investigates several possible scenarios (e.g., using more sophisticated models to predict future costs) is necessary before any conclusive evidence can be obtained on whether optimal marketing costs provide better targeting over regression-based estimates of marketing costs. In summary, the results indicate that selecting customers on the basis of their expected (future) marketing costs leads to better targeting of profitable customers than selecting customers on the basis of their historic marketing costs.

IMPLICATIONS

Our academic–practitioner collaboration revealed major concerns among practitioners regarding the implementation of CLV for customer selection, for which there are no explicit recommendations in the literature. First, given the

[6]The cost of serving customers refers to the cost of the marketing communication directed toward the customers.

uncertainty inherent in predicting customer behavior, managers are wary about taking actions based on CLV that is calculated using predictions of customer behavior over the long run. Second, there is also no guidance on a forward-looking, individual, customer-level cost allocation rule when computing CLV. These concerns formed the motivation for our study. We proposed a Bayesian decision theory–based customer selection framework, which explicitly accounts for uncertainty in predicting customer behavior over the long run, for selecting customers using CLV. We proposed a joint model for predicting purchase timing and quantity that enables us to compute CLV accurately and is amenable to the proposed selection procedure. The major recommendations our study provides for managers interested in customer selection are as follows:

- The proposed joint purchase timing–quantity model has better predictive accuracy than other currently available model frameworks, thus emphasizing the need to account for the dependence between purchase timing and quantity.
- Selecting customers using a Bayesian decision theory–based customer selection framework leads to better identification of profitable customers.
- During the customer selection process, in addition to quantity, managers should focus on the expected costs of serving a customer. The optimal costs derived using estimates of customers' historic responsiveness to marketing communication provide a good estimate of the expected costs of serving a customer and aids in identifying profitable customers.

The proposed Bayesian decision theory–based selection strategy identifies profitable customers better than current practices at the collaborating firm. The firm had contacted all the customers in the analysis sample until they purchased or the budget was exhausted. So the firm's current CRM policy forms a natural control group for our analysis. Although the proposed Bayesian decision theory–based strategy is better at identifying profitable customers, the level of profits provided by the customers could be higher if the firm had used the recommended optimal marketing decision variables.

The purchase timing–quantity model we proposed herein is also suitable for computation of several other commonly used metrics for customer selection. For example, recency, frequency, and monetary value (RFM) is a commonly used metric for customer selection in direct marketing. In general, the RFM metric is represented as

$$\text{RFM}_i = w_r \times R_i + w_f \times F_i + w_m \times \text{CM}_i, \tag{14}$$

where

RFM_i = the RFM score for customer i,
R_i = the predicted recency of purchase for customer i,
F_i = the predicted frequency of purchase for customer i,
CM_i = the predicted monetary value of purchase for customer i, and
w_r, w_f, w_m = factor scores obtained from a factor analysis of RFM values of all the customers. We calculated RFM using prior data.

The RFM value projections for the future periods can be obtained from predictions of the purchase timing–quantity model as follows:

$$\text{Recency} = R_i = 12 - \sum_{j=T^*+1}^{T^*+T_i} \hat{t}_{i,j} - T_{i,lc}, \tag{15a}$$

$$\text{Frequency} = F = \text{ratio of 12 to } \hat{t}_i\text{, and} \tag{15b}$$

$$\text{Contribution margin} = \text{CM}_i = \sum_{j=T^*+1}^{T^*+T_i} \hat{Q}_{i,j} \times M, \tag{15c}$$

where

$\hat{t}_i$ = the expected interpurchase time for customer i, which is obtained from the expected value of the purchase timing model;
T_i = the number of purchases made by customer i in one year; and
$T_{i,lc}$ = the period of the last purchase made by customer i in the calibration sample.

Guidelines for Implementation

A firm can implement the proposed strategy with the hardware systems it currently uses for collecting customer information. The one-time software development cost for the proposed model would not exceed $100,000. With regard to time required, the model proposed here can be implemented within three months, which is a reasonable time frame for a customer selection process in B2B settings. Given the potential for improvement in the NPV of approximately $784,000 (difference in the NPV of profits between Cells B and D in Table 4 for the sample under study), the benefits of using the suggested selection strategy outweigh the costs. Suppose that a firm intends to implement the proposed customer selection process for a larger customer segment (e.g., 100,000 customers); then, it is important to focus on the following:

1. *Plan for a longer period for model estimation*. An attractive property of the proposed joint purchase timing–quantity model is the accommodation of customer heterogeneity through the segment formulation (Equations 6 and 7), the purchase rate parameter (λ_{ij}, Equation 3) in the purchase timing component, and the intercept ($\delta_{0,i}$) in the purchase quantity component (Equation 5). However, allowing for heterogeneity also increases the estimation time and the time required to score customers. Therefore, managers should allow for sufficient time for estimation of the parameters and the scoring of the customers. Current developments in computing technology reduce the time required to estimate complex models, but still, the additional time required to estimate the proposed model is not negligible compared with other aggregate models, such as the ESP model estimated in the collaborating firm.
2. *Align model reestimation cycles with the resource allocation cycles*. At each reestimation cycle, new data obtained from customer responses to marketing campaigns in the recent past should be included. When including new data for model reestimation, a rolling-window approach can be adopted for the calibration sample to control for sample size explosion. For example, managers can use a three-year rolling window for the reestimation sample. If the model is reestimated every year, information on customer transactions four years before can be dropped from the model estimation sample when new data are obtained for the most recent year. Proper sensitivity analysis of the model predictions should be carried out before deciding on the appropriate time frame for the rolling window.

3. *Use an optimal level of marketing contacts as a benchmark*. Because the model and, thus, the optimal level of marketing contacts do not include several factors, such as the intensity of marketing contacts from the competition, firms should plan for a slightly higher frequency of customer contacts than the suggested optimal level of contacts. Planning for a higher-than-optimal frequency of marketing contacts would allow firms to keep contacting the nonresponsive customers even when the suggested optimal level of marketing contacts is reached.
4. *Incorporate salesperson intuitions*. For example, consultations with salespeople would be valuable in determining the extent to which the planned contact frequency should be higher than the suggested optimal contact frequency.
5. *Evaluate model reformulation*. Substantial changes in the market or environmental factors should determine the appropriate time window for model reformulation.

Enhancing Marketing Productivity

Several researchers have proposed that customer profits can be substantially improved if managers customize marketing contacts to individual customer preferences (Ansari and Mela 2003; Reinartz, Thomas, and Kumar 2005; Venkatesan and Kumar 2004). However, practitioner implementation of this recommendation is limited because the previous academic research does not show any causal link between using optimal contact levels and maximized customer profits. Given the huge opportunity costs of conducting a field experiment to establish a causal link between optimal contact levels and maximized customer profits, we adopt an alternative approach to establish the benefits from harnessing the power of historic customer information to improve marketing productivity. Our results show that deriving optimal contact levels can lead to improved targeting of profitable customers and, thus, to improved marketing productivity. In addition, the optimal contact levels are a better estimate of future costs than historic marketing costs. Although a causal link between optimal marketing contacts and maximized customer profits is still not established, we believe that the results from our selection exercise improve the adoption of customization of marketing contacts among practitioners at least for the purposes of customer targeting. The results also imply that a proactive customer selection strategy will enable managers to target profitable customers, leading to higher profits given a fixed marketing budget.

Estimating Returns from Marketing Actions

Our proposed strategy enables managers to (1) identify which customers would be more profitable in the future and (2) estimate the expected costs of serving the selected customers. These results and the selection exercise would enable managers to justify investments in customized marketing for retained customers and to provide a better estimate of the budget that would be required for achieving their goals. Managers can use the proposed CLV framework and the corresponding optimal marketing costs that maximize CLV to show a link between marketing investments and the returns expected from each customer. Marketing managers can use the selection exercise we propose to show top management the return on investment that can be obtained from a marketing strategy that targets individual customers. In addition, the CLV measure and the corresponding optimal marketing costs can form the basis for discussions regarding marketing budgets over the long run.

LIMITATIONS AND FURTHER RESEARCH

The study has limitations that can be addressed by future studies. The results of this study are from a sample of customers in the high-technology industry. Further replications are necessary across samples within the sponsoring firm before the findings can be generalized to the entire population of customers in this firm. In addition, further research should investigate whether the results are generalizable to other industries and settings. A fertile area for further research would be to evaluate the value of information regarding the competitor's marketing actions on the forecasts of customer profits. We consider only the average levels of optimal communication strategy in each channel. However, organizations can further improve the efficiency of their customer selection process by deriving the optimal sequence of customer contacts across different channels.

We do not explicitly accommodate the firm's CRM strategy in the purchase timing–quantity model. We expect that doing so would improve both the accuracy of the parameter estimates of the model and the accuracy of the CLV measure and therefore would provide greater support for our use of the proposed Bayesian decision theory–based procedure for customer selection. As we explained previously, the collaborating firm used ESP to decide which customers to contact. However, after a customer is selected for contact, there is no specific guidance for the level of contacts. In general, the customers are contacted until they make a purchase. In our analysis, we included only customers who were contacted in the past because we wanted to understand how the collaborating firm could improve the efficiency of its current process. Because the firm does not follow any particular contact strategy for the customers who were selected, we believe that the extent of bias in the parameter estimates of the proposed model would be minimal in our context. Although our model framework represents the demand side of customer responses given firm contacts, further research could extend our framework to model the supply side of a firm's policies of determining the level of contacts (see, e.g., Manchanda, Rossi, and Chintagunta 2004).

Our study indicates that in addition to predicting future customer profits, estimating a customer's future costs contributes to better selection capabilities. Further research is necessary to identify frameworks for planning the future marketing communications for a customer who is dynamically updated according to his or her responses. Finally, the question that arises from our analyses is how the recommendations from an optimization framework would work when implemented in the real market. Although our study is a step in the right direction to assess the accountability of marketing actions, a field experiment that tests the recommendations of such a framework on a test group against a control group that is managed according to existing norms would provide a stronger justification for CRM-based efforts.

REFERENCES

Allenby, Greg M., Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94 (June), 365–74.

Ansari, Asim and Carl Mela (2003), "E-Customization," *Journal of Marketing Research*, 40 (May), 131–45.

Armstrong, Scott, Vicki Morvitz, and V. Kumar (2000), "Sales Forecasts for Existing Consumer Products and Services: Do Purchase Intentions Contribute to Accuracy?" *International Journal of Forecasting*, 16, 383–97.

Baltagi, Badi H. (1998), *Econometric Analysis of Panel Data*. New York: John Wiley & Sons.

Boatwright, Peter, Sharad Borle, and Joseph B. Kadane (2003), "A Model of the Joint Distribution of Purchase Quantity and Timing," *Journal of the American Statistical Association*, 98 (463), 564–72.

Bolton, Ruth N., Katherine Lemon, and Peter C. Verhoef, (2004), "The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research," *Journal of the Academy of Marketing Science*, 32 (3) 271–92.

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

Bult, Jan Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378-–394.

Chintagunta, Pradeep (1993), "Investigating the Purchase Incidence, Brand Choice, and Purchase Quantity Decisions of Households," *Marketing Science*, 12 (2), 184–208.

Fournier, Susan, Susan Dobscha, and David Glen Mick (1997), "Preventing the Premature Death of Relationship Marketing," *Harvard Business Review*, 75 (January–February), 2–8.

Gelfand, A.E. and Dipak K. Dey (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of Royal Statistical Society, Series B*, 56 (3), 501–514.

Grewal, Rajdeep, James M. Corner, and Raj Mehta, (2001), "An Investigation into the Antecedents of Organizational Participation in Business-to-Business Electronic Markets," *Journal of Marketing*, 65 (July), 17–34.

Krishnamurthi, Lakshman and S.P. Raj (1988), "A Model of Brand Choice and Purchase Quantity Price Sensitivities," *Marketing Science*, 7 (1), 1–20.

Manchanda, Puneet, Peter E. Rossi, and Pradeep Chintagunta (2004), "Response Modeling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41 (November), 467–78.

McDonald, James B. and Richard J. Butler (1990), "Regression Models for Positive Random Variables," *Journal of Econometrics*, 43, 227–51.

Mohr, Jakki and John R. Nevin (1990), "Communication Strategies in Marketing Channels: A Theoretical Perspective," *Journal of Marketing*, 54 (October), 36–51.

Morgan, Robert M. and Shelby D. Hunt (1994), "The Commitment–Trust Theory of Relationship Marketing," *Journal of Marketing*, 58 (July), 20–38.

Niraj, Rakesh, Mahendra Gupta, and Chakravarthi Narasimhan (2001), "Customer Profitability in a Supply Chain," *Journal of Marketing*, 65 (July), 1–16.

Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17–35.

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

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

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304–328.

———, T. Ambler, G. Carpenter, V. Kumar, and R. Srivastava (2004), "Measuring Marketing Productivity: Current Knowledge and Future Directions," *Journal of Marketing*, 68 (October), 76–89.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml, (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (January), 109–128.

Schmittlein, David, Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (January), 1–24.

Schultz, Don E. (2003), "Marketing Communications Allocation: Asking the Right Questions," in *Measuring and Allocating Marcom Budgets: Seven Expert Points of View*, Rajeev Batra and David Reibstein, eds. Cambridge, MA: Marketing Science Institute, 37–44.

Steenburg, Thomas J., Andrew Ainslie, and Peder Hans Engerbretson (2003), "Massively Categorical Variables: Revealing the Information in Zip Codes," *Marketing Science*, 22 (1), 40-57.

Venkatesan, Rajkumar, T.V. Krishnan, and V. Kumar (2004), "Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares," *Marketing Science*, 23 (3), 451–64.

——— and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (October), 106–25.

Webster, Frederick E., Jr. (1992), "The Changing Role of Marketing in the Corporation," *Journal of Marketing*, 56 (October), 1–17.