MICHAEL BRAUN, DAVID A. SCHWEIDEL, and ELI STEIN*

Dynamic customer targeting is a common task for marketers actively managing customer relationships. Such efforts can be guided by insight into the return on investment from marketing interventions, which can be derived as the increase in the present value of a customer's expected future transactions. Using the popular latent attrition framework, one could estimate this value by manipulating the levels of a set of nonstationary covariates. The authors propose such a model that incorporates transaction-specific attributes and maintains standard assumptions of unobserved heterogeneity. They demonstrate how firms can approximate an upper bound on the appropriate amount to invest in retaining a customer and demonstrate that this amount depends on customers' past purchase activity—namely, the recency and frequency of past customer purchases. Using data from a business-to-business service provider as their empirical application, the authors apply the model to estimate the revenue the service provider loses when it fails to deliver a customer's requested level of service. They also show that the lost revenue is larger than the corresponding expected gain that would result from exceeding a customer's requested level of service. The authors discuss the implications of their findings for marketers in terms of managing customer relationships.

Keywords: services marketing, customer retention, probability models, marketing return on investment, customer value

Online Supplement: http://dx.doi.org/10.1509/jmr.13.0377

Transaction Attributes and Customer Valuation

According to IBM's recent survey of chief marketing officers, 63% of respondents said that return on investment (ROI) would be the most important measure of success in the next three to five years (IBM 2011). Braun and Schweidel (2011) argue that marketing ROI should be measured in terms of an expected change in the residual value of the customer base that occurs from a marketing intervention. Latent attrition models, such as the Pareto/negative binomial distribution (Pareto/NBD; Schmittlein, Morrison, and Colombo 1987) and the beta geometric/negative binomial distribution (BG/NBD; Fader, Hardie, and Lee 2005a), are useful for estimating residual value, but they do not lend themselves easily to incorporating covariates that change from transaction to transaction. In the marketing ROI domain, possible examples of these covariates include attributes of the marketing mix, exogenous conditions at the time of the transaction (e.g., weather effects), or an investment in improved customer experiences.

Research has found that the quality of customer information to which a firm has access, which encompasses being both broad and up-to-date, moderates the profitability of customer prioritization (Homburg, Droll, and Totzek 2008). Similarly, Mithas, Krishnan, and Fornell (2005) posit that the value of customer relationship management tools lies in their ability to facilitate learning about customers over the course of multiple interactions, the insights from which can then be used to target customers dynamically with tailored offerings. Although some researchers have proposed methods of incorporating both time-invariant and time-variant predictors into customer base analyses (Abe 2009; Schweidel and Knox 2013), these models stop short of establishing the link between the

*Michael Braun is Associate Professor of Marketing, Cox School of Business, Southern Methodist University (e-mail: braunm@smu.edu). David A. Schweidel is Associate Professor of Marketing, Goizueta Business School, Emory University (e-mail: dschweidel@emory.edu). Eli Stein is an alumnus of Harvard College, Harvard University (e-mail: estein@post.harvard.edu). Marnik Dekimpe served as associate editor for this article.

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

Journal of Marketing Research
Vol. LII (December 2015), 848–864
DOI: 10.1509/jmr.13.0377

effects of transaction-specific attributes and forecasts of residual value. Thus, how to use information about transaction attributes to assess the increased customer value stemming from marketing efforts remains an open and important area of research.

In this article, we propose a latent attrition model that integrates transaction attributes into a probability model of customer retention and lifetime value. One application of this model (which we use in our empirical analysis) is to evaluate the impact of a customer's service experience on customer retention. Consider the following scenario. A local business (i.e., the customer) contracts with a service provider to meet a recurring business need, such as copywriting services. Drawing from examples of the provider's work, the customer forms some expectation of the provider's caliber and timeliness. If the provider delivers a top-notch experience, the customer may engage the provider the next time (s)he needs similar services. By increasing the retention probability for that customer, all else being equal, the service experience delivered by the provider's team may have generated additional revenue in subsequent time periods. In contrast, if the provider falls short and delivers a subpar experience, the customer may be more likely to search for another provider when the need for similar services arises in the future. When this happens, the future revenue stream from the customer falls to zero. In this example, the attributes of a single customer–firm interaction could affect the future revenue stream. The "transaction attribute" in this case is an indicator of the customer's service experience. More generally, transaction attributes may refer to some aspect of the marketing mix, the employee who processed a customer's transaction, or any information about a customer's transaction that the firm has collected.

Although there is an intuitive relationship between transaction attributes and retention probabilities and, thus, customer value, extant empirical research often has failed to differentiate transactions other than with regard to the times at which the events occur. Models such as the Pareto/NBD and the BG/NBD rely on summary statistics of prior transactional activity—namely, the total number of transactions (frequency) and the time of the most recent transaction (recency). However, by aggregating the data to this level, we lose information about other characteristics of each transaction. Take the case of two customers with the same number of transactions during the last year, with their most recent transaction occurring on the same day. If we ignore attributes of those transactions, these two customers' transaction histories (and the resulting predictions of residual value) are identical. If a firm had access to additional information about the last customer–firm touch point, it could generate different predictions of what each customer would do in the future. Failing to meet stated and established standards of service, for example, may lead the firm to believe that a customer is now at a greater risk for churn, whereas exceeding these standards may lead to perceptions that this risk is lower (Ho, Park, and Zhou 2006).

Exploiting information about the attributes of each transaction gives firms additional guidance for managing customer relationships compared with the information provided by frequency and recency alone. The value of transaction attribute data comes from how that information affects the firm's decisions. We assess the effect of a particular transaction attribute in terms of the projected change in discounted expected residual transactions (DERT; Fader, Hardie, and Lee 2005b; Fader, Hardie, and Shang 2010). Discounted expected residual transactions are the present values of expected future transactions, taking into account the probability that a customer may have already churned or will churn in the future. Differences in the transaction attributes will produce variation in DERT beyond that which is captured by the recency and frequency of transactions. Attribute data can affect firm decisions in two ways. First, ignoring attributes can lead to different estimates of DERT, which in turn may lead to suboptimal decisions based on bad information. Second, the "incremental DERT" that comes from manipulating the transaction attributes (e.g., by increasing investment in retaining a specific customer immediately before a transaction) can serve as an upper limit for such a marketing investment (Braun and Schweidel 2011). For example, if a firm could estimate the effect of a marketing-mix variable on churn, the incremental DERT would be the difference between the forecast of DERT under the marketing treatment and a baseline DERT without it. To the best of our knowledge, our research is the first to examine the value of information about transaction attributes in terms of the change in expected future customer transactions.

In the next section, we provide a brief discussion of relevant literature in the customer base analysis and service quality areas, the latter of which relates to the context of our empirical example. We show how our research adds to the associated body of knowledge. Then, we describe the model in terms of a likelihood function and derive both prior and posterior DERT that can be expressed either in closed form or as a summation. The subsequent empirical analysis, which employs a data set from a noncontractual service provider, illustrates a case in which including transaction attribute data adds predictive power to the model. Finally, we show that incremental DERT has a nonlinear and nonmonotonic relationship with customer recency and frequency. The patterns suggest that falling short of the requested service level has a smaller effect on retaining customers who are either likely to have already churned or highly *unlikely* to have churned, compared with the effect on customers for whom there is more uncertainty in their active status.

RELATED LITERATURE

Latent attrition models such as the Pareto/NBD and the BG/NBD are workhorse models of customer base analysis (Fader, Hardie, and Lee 2005a, b; Schmittlein, Morrison, and Colombo 1987). A notable limitation of this class of models is the difficulty of incorporating information about attributes that accompany each customer–firm interaction, such as marketing actions that vary across transactions. Recognizing this gap in the literature, Ho, Park, and Zhou (2006) propose incorporating customer satisfaction into the latent attrition framework. Their model assumes that customer satisfaction affects the rate at which customers conduct transactions, and they demonstrate how satisfaction can affect the attrition process. Although Ho, Park, and Zhou do consider information that is specific to the customer–firm interaction (i.e., satisfaction), their model is analytic as opposed to empirical. Nevertheless, they illustrate the importance of incorporating customer satisfaction and, more broadly, customer–firm interaction information, into estimates of customer value.

Another notable difference between the Ho, Park, and Zhou (2006) model and empirical latent attrition models is that Ho, Park, and Zhou assume homogeneous purchase and attrition processes. In addition to capturing variation across customers, models that allow for unobserved heterogeneity let firms update their expectations of customer behavior as new data become available. These posterior inferences are necessary both for valuing customers (Braun and Schweidel 2011; Fader, Hardie, and Lee 2005a) and for assessing the impact of marketing efforts. Schweidel and Knox (2013) illustrate this idea with a joint model of people's donation activity and the direct marketing efforts of a nonprofit organization, accounting for the potentially nonrandom nature of marketing efforts. In their example, the authors allow for direct marketing activity to affect the likelihood of donation each month, the amount of a donation conditional on the donation occurring, and the likelihood with which a donor becomes inactive. To account for unobserved heterogeneity, Schweidel and Knox apply a latent class structure. Although their model allows marketing actions to influence each of the processes that may affect customer value, they do not consider how a person's transaction history may affect expectations of future activity. Moreover, they do not consider how their framework could be adapted to estimate measures such as expectations of future purchases, customer lifetime value, or residual value.

Like Schweidel and Knox (2013), Knox and Van Oest (2014) also employ a latent class structure to account for heterogeneity across customers in their investigation of the impact of customer complaints on customer churn. They assess the impact of customer complaints and recovery by the firm for two types of customers: a new customer and an established customer. The authors demonstrate that the residual value of customers following a complaint varies with both the customers' previous purchase activity and previous complaints. Consistent with research that has investigated the profitability of behavior-based marketing actions, Knox and Van Oest distinguish between the effects of marketing interventions on new and established customers (Pazgal and Soberman 2008; Shin and Sudhir 2010; Villas-Boas 1999). Although extant work has documented the benefits of differentiating new and established customers, such work often does not aim to provide insight into how marketing interventions may affect established customers with different transaction histories. We contribute to this stream of research by developing a modeling framework that allows us to conduct a systematic investigation into how customers' recency and frequency of past transactions (Fader, Hardie, and Lee 2005b) affect the incremental impact of marketing efforts, which can enable marketers to target customers with increased precision.

Although there are many different attributes a transaction could possess, our empirical analysis is in the domain of service quality. Several researchers have studied service quality and its relationship to customer expectations. Boulding et al. (1993) find that customers' evaluations of a service encounter are affected by both their prior expectations of what will and should occur and the quality of service delivered on recent service encounters. In essence, "will" and "should" expectations for a service encounter are a weighted average of prior expectations and the recently experienced service. Boulding, Kalra, and Staelin (1999) further investigate the process by which expectations are updated. In addition to affecting a customer's cumulative opinion, the authors find that prior beliefs also affect how experiences are viewed. As a result, prior expectations deliver a "double whammy" to evaluations of quality. This suggests that all service encounters are not equal in the eyes of consumers, because their previous experiences affect the way they view service encounters. For example, the exact same level of quality might exceed expectations in a midrange family restaurant but fail to meet expectations in a fancy bistro. Yet extant customer valuation models in both noncontractual and contractual settings often assume that the "touch points" associated with customer–firm interactions are equivalent to each other.

Rust et al. (1999) further investigate the role of customer expectations in perceptions of quality. Rather than focusing on the average expectation across customers, the authors highlight the importance of the distribution of customer expectations. They tackle several myths regarding the level of service that providers should deliver to their customers. In contrast to the popularly held belief that firms must exceed expectations, the authors find evidence that simply meeting customers' expectations can result in a positive shift in preferences. They also find that service encounters that are slightly below expectations may not affect customers' preferences at all. Though these results are provocative, the authors recognize that additional research is needed because they conducted their investigation in a laboratory setting and relied on self-reports.

In addition to the work that has been conducted on service quality, our research is also related to the customer satisfaction literature. Bolton (1998) investigates the impact of customer satisfaction on the duration for which customers continue to subscribe to a contractual service. She finds that reported customer satisfaction with the service, solicited before the customer decides whether to remain a subscriber or cancel service, is positively related to the duration for which a customer will retain service. She also finds evidence that recent experiences with the service provider are weighed differently depending on whether the customer evaluated the experience as positive or negative. To the best of our knowledge, research on customer valuation has not incorporated this differential weighting of customer experiences into estimates of customers' future behavior.

MODEL

Model Derivation

In this section, we propose a general form of a latent attrition model that incorporates transaction attributes. To keep terminology consistent with the empirical example in the following section, we say that the customer of the firm places orders for jobs, and the firm fills those orders by completing the jobs. Thus, orders and jobs always occur in a pair and are indexed by k. We assume that these jobs are completed the instant the order is placed, so we index calendar time for orders and jobs by t. Without loss of generality, we define a unit of calendar time as one week. The service was introduced to the marketplace at time $t = 0$, and T is the week of the end of the observation period. Let t_1 be the week of the customer's first order; let x be the number of orders between times t_1 and T, including that first order at t_1; and let t_k be the time of order k. Therefore, t_x is the order time of the final observed job. For clarity, we suppress the customer-specific indices on t and x in the model exposition.

Our baseline model is a variant of the BG/NBD model for noncontractual customer base analysis (Fader, Hardie, and Lee 2005a). Immediately before the customer places an initial order at time t_1, (s)he is in an active state. While active, the customer places orders according to a Poisson process with rate λ. After each job (including the first one), a customer may churn, resulting in that order being last. With probability p_k, the customer churns after the kth job and transitions from the active state to the inactive state. When this occurs, we assume that the customer is lost for good and will not place any more orders. If the customer does not churn, the time until the next order, $t_{k+1} - t_k$, is a realization of an exponential random variable with rate λ. We never observe directly when, or if, a customer churns, although if a customer places x orders, (s)he must have survived $x - 1$ possible churn opportunities.

For a customer who places x orders between times t_1 and T, the joint density of the $x - 1$ interorder times is the product of $x - 1$ exponential densities. For this customer, there could not have been any orders between times t_x and T. This "hiatus" could occur in one of two ways. One possibility is that the customer may have churned after job x, with probability p_x. Alternatively, the customer may have "survived," with probability $1 - p_x$, but the time of the next order would be sometime after T. Thus, conditional on surviving x jobs, the probability of not observing any more jobs before time T is $e^{-\lambda(T-t_x)}$. Thus, the conditional data likelihood for a single customer is

$$(1)\quad f(x, t_{2:x}|\lambda, p_{1:x}) = \lambda^{x-1}e^{-\lambda(t_x-t_1)}\left[\prod_{k=1}^{x-1}(1-p_k)\right]\left[p_x + (1-p_x)e^{-\lambda(T-t_x)}\right].$$

If p_k were time invariant (i.e., the same for all k), Equation 1 would be the individual-level likelihood in the BG/NBD.[1] To incorporate transaction-specific information, we allow p_k to vary across orders in our model. We define the probability of becoming inactive by transitioning to the inactive state immediately after job k as $p_k = 1 - e^{-\theta q_k}$ and define $B_k = \sum_{j=1}^{k} q_k$, where q_k is a nonnegative scalar value that can influence the probability that a customer transitions to the inactive state after job k. If we restrict $q_k = 1$ for all k, then $p = 1 - e^{-\theta}$ or, alternatively, $\theta = -\log(1 - p)$. The expression q_k—and, thus, B_k—could be a function of further parameters and observed data, such as the transaction attributes. For example, we might give q_k a log-linear structure, where $\log q_k = \beta' z_k$, β is a vector of homogeneous coefficients, and z_k is a vector of covariates that represents attributes of transaction k. Substituting these definitions into Equation 1, we have

$$(2)\quad f(x, t_{2:x}|\lambda, \theta, q_{1:x}) = \lambda^{x-1}e^{-\lambda(t_x-t_1)-\theta B_{x-1}}\left[1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)\right].$$

The expression of likelihood in Equation 2 assumes that all customers place orders at the same rate and that all customers have the same baseline propensity to churn after each job. To incorporate heterogeneity of latent characteristics into the model, we let λ and θ vary across the population according to gamma distributions, where $\lambda \sim \mathcal{G}_\lambda(r, a)$ and $\theta \sim \mathcal{G}_\theta(s, b)$. Integrating over these latent parameters, we get the following marginal likelihood:

$$(3)\quad \mathcal{L} = \frac{\Gamma(r+x-1)}{\Gamma(r)}\frac{a^r}{(a+t_x-t_1)^{r+x-1}}\left(\frac{b}{b+B_{x-1}}\right)^s \times\left\{1 - \left(\frac{b+B_{x-1}}{b+B_x}\right)^s\left[1 - \left(\frac{a+t_x-t_1}{a+T-t_1}\right)^{r+x-1}\right]\right\}.$$

Equation A1 in the Appendix provides a detailed derivation of the marginal likelihood. Grassia (1977) discusses the issue of transforming a gamma-distributed random variable to yield a value between 0 and 1. If p_k were constant over time and varied across the population according to a beta distribution, the marginal likelihood would be the same as the BG/NBD. Griffiths and Schafer (1981) show that Grassia's method and a beta distribution are "practically identical," and that choice between them could be based "entirely on mathematical convenience" (p. 245). Our approach lets us estimate model parameters using standard maximum likelihood techniques even when the attrition probability depends on transaction-specific covariates.

Certain design decisions enable us to maintain some degree of computational efficiency. We allow for unobserved heterogeneity in λ and θ by carefully choosing a parametric family of independent mixing distributions. The B_x term is heterogeneous across observable characteristics that vary across both individuals and time. We also allow for unobserved nonstationarity in some of the parameters in B_x (see Equations 8 and 9).[2]

Conditional Expectations and DERT

After a manager has parameter estimates in hand, (s)he might be interested in the number of orders we might receive from a newly acquired customer during a period of t weeks. In the Appendix, we show that the prior expected value of this order count is

$$(4)\quad E[X(t)] = \sum_{k=1}^{\infty}\left(\frac{b}{b+B_k}\right)^s\tilde{\mathbb{B}}\left(\frac{t}{a+t}; k, r\right).$$

Therefore, the manager can estimate the expected number of orders by truncating this infinite series. The function

[1]As with the BG/NBD, in our model a high transaction rate suggests additional attrition opportunities.

[2]Among the alternative specifications considered was one in which we allow for heterogeneity in β with a latent class structure. The estimated probability of being in the first latent class was $p = 1$, suggesting that the additional model complexity is not warranted. We also estimated a model that allows λ to vary with z_k. The Hessian in the resulting model was singular, so the parameters could not be identified when we assumed covariates to affect both the transaction rate and the attrition process simultaneously.

$\tilde{\mathbb{B}}(x; a, b)$ is the regularized beta function, which also happens to be the cumulative distribution function (CDF) of a beta distribution, with parameters a and b, evaluated at x.[3]

One way to interpret the summation in Equation 4 is as the sum of the probabilities of ordering k jobs before time t for all possible values of k. These jobs are hypothetical, so we need a model for each q_k that composes B_k, which is the cumulative sum of q_1, q_2, ..., q_k. In general, one could simulate multiple sequences of q_k from that model, truncated at a sufficiently large value of k, and then average E[X(t)] across sequences. An alternative heuristic is to replace each B_k with its mean. This approximation will be most accurate when the variance in B_k is very small, which, as we show subsequently, is the case in our empirical application. This approximation is not needed to estimate the model itself but only to calculate the expected number of transactions without resorting to the use of simulations.

A conceptually useful expression is the probability that a customer is still active at time T.

$$\text{(5)} \qquad P(\mathcal{A}) = \left\{1 - \left(\frac{a + T - t_1}{a + t_x - t_1}\right)^{r+x-1}\left[1 - \left(\frac{b + B_x}{b + B_{x-1}}\right)^s\right]\right\}^{-1}.$$

For the derivation for Equation 5, see the Appendix.

A manager might also want to know how many orders (s)he can expect from an existing customer during the next t^* periods, given an observed transaction history. In the Appendix, we show that this posterior expected number of future transactions is

$$\text{(6)} \qquad E^*\left[X\left(t^*\right)\middle|x, t_x\right] = P(\mathcal{A})\sum_{k=1}^{\infty}\left(\frac{B_x + b}{B_x + B_k + b}\right)^s \tilde{\mathbb{B}}\left(\frac{t^*}{t^* + a + T - t_1}; k, r + x - 1\right).$$

In Equation 6, the index of the summation k refers to the potential orders that are made after time T. As discussed previously, we can either model B_k so that we may simulate future values of B_k explicitly or substitute $E(B_k)$ as an approximation. We present the prior and posterior probability mass functions for the number of orders (i.e., to express the probability of placing a particular number of orders during some future number of weeks) in the Web Appendix.

Although it is useful to know the expected number of future orders, orders are placed at different times. One order may be placed at time T + 1, and the next order may not be placed until well into the future. Given the time value of money, orders that occur sooner are more valuable than orders that are placed later. Therefore, an appropriate metric for the expected number of a customer's future transactions should discount those transactions back to the present. The value for DERT is proportional to a customer's residual lifetime value when the margin is constant (Fader, Hardie, and Shang 2010). Let δ be a discount factor that captures the time value of money, so a dollar earned t weeks from now is worth δ^t today (for notational simplicity, we reset the counter of t so that t = 0 at T, and we assume that payments are made at the end of the week). The posterior estimate for the DERT of this customer is the sum of discounted incremental expected orders:

$$\text{(7)} \qquad \begin{aligned} \text{DERT} &= \sum_{t=1}^{\infty}\left(E^*[X(t)|x, t_x, B_x, T] - E^*[X(t-1)|x, t_x, B_x, T]\right)\delta^t \\ &= P(\mathcal{A})\sum_{k=1}^{\infty}\left(\frac{B_x + b}{B_x + B_k + b}\right)^s \\ &\quad \times \sum_{t=1}^{\infty}\delta^t\left[\tilde{\mathbb{B}}\left(\frac{t}{a + T - t_1 + t}; k, r + x - 1\right) - \tilde{\mathbb{B}}\left(\frac{t-1}{a + T - t_1 + t - 1}; k, r + x - 1\right)\right] \\ &= (1 - \delta)P(\mathcal{A})\sum_{k=1}^{\infty}\left(\frac{B_x + b}{B_x + B_k + b}\right)^s \\ &\quad \times \sum_{t=1}^{\infty}\delta^t\tilde{\mathbb{B}}\left(\frac{t}{a + T - t_1 + t}; k, r + x - 1\right). \end{aligned}$$

These future transactions depend on several different elements. The parameters r, a, s, and b capture the distribution of order rates and baseline churn likelihoods across the population (e.g., for any randomly chosen member of the population, $E(\lambda) = r/a$ and $E(\theta) = s/b$). Through $P(\mathcal{A})$, customers with low t_x and high x might be more likely to have already become inactive, so there is a low probability that these customers will conduct transactions in the future. Customers with high x and high t_x are more likely to be "alive" and to order often, so their DERT should be high.

Like all statistical models, this model is intended as a schematic of the actual data-generating process. To give the model some useful parametric structure, we treat the latent attrition process as a manifestation of a random variable. Although more complicated versions of the model could be proposed (e.g., a model that allows for duration dependence in purchase times or contagion across customers in their propensities to churn), we favor parsimony so as to avoid overparameterizing the model given certain limitations in a typical transactional data set.

EMPIRICAL ANALYSIS

The context in which we study the role of quality on customers' future transactions is that of an online market for freelance writing services. The firm in question operates a website on which customers can post orders for "jobs," and from which writers can claim jobs to complete. The types of jobs vary greatly. One example would be a 100-word description of a product that the customer, an online retailer, is selling on its website. Another is a 500-word summary of what participants at a conference might do for fun when exploring the host city. Orders include all of the information a writer would need to complete the job: the topic area (e.g., sports, health), intended audience, word count, and so forth. Customers are encouraged to be as specific as possible in their requirements to ensure that they will be satisfied with the results. In our taxonomy, we consider an order to be equivalent to the posting of a job.

Customers also choose a minimum rating, or grade, for the writers who are eligible to claim the order. The firm maintains a bank of reviewers who screen and rate the writers who register with the website. These reviewers are employed directly by the firm and are considered experts in

[3]Appendix Table A1 provides a glossary of many of the functions we use in this article.

evaluating prose (many have master of arts degrees or similar qualifications). Upon initial application, a writer submits a writing sample, and a reviewer rates the writer as A, B, C, or D. The firm's website provides examples of work from the different rating categories, so customers have a general idea about the differences to expect across the ratings. Ratings differ according to objective criteria such as accuracy, grammar, style, and vocabulary. A D-rated writer might produce work with errors and simple sentence structure with no creative insight, whereas work from an A-level writer will be of professional quality.

Customers pay, and writers earn, on a per-word basis, wherein the charge for each word depends on the rating in the order. The firm claims a fixed percentage of this fee, plus a small (less than a dollar) charge per order. Writers claim jobs from a list on a first-claim basis, so there is no bidding involved. Writers may claim jobs that are rated below their own ratings (e.g., an A-rated writer can choose a project from any level, but a B-rated writer cannot choose an A-rated order). In such cases, writers are paid the lower per-word fee. The company has told us that it has not experienced shortages of writers, with most job specifications being claimed within a day. Writers have another day to complete the job, and nearly all jobs are completed within 24 hours of posting.

Sometime after the writer returns the completed job to the customer, the firm's bank of reviewers assigns each job a grade. Customers are not involved in this grading process, and neither customers nor writers ever see the grade for a particular job. However, a writer's rating can be adjusted according to his or her grade history. This gives the writer an incentive to complete the job well; the grades determine whether the writer's rating is adjusted up or down. The reviewers try to rate jobs as accurately and objectively as possible to ensure that customers receive the quality they pay for and to reclassify writers as necessary. Writers can only be elevated to the A-level manually, so the firm classifies all A-rated and B-rated jobs together in an A/B class. Reviewers may also assign a grade of E for completed jobs that do not meet even minimum standards.

Data Summary

Our master data set includes all completed jobs from the launch of the company in June 2008 to the end of our observation period at the end of July 2011. We restrict our analysis to customers in either the United States or Canada whose first order takes place before the end of 2010 and to jobs for which the language is English. This data set contains information on 24,059 completed jobs ordered by 3,048 distinct customers. For each job, we have identifiers for the customer and writer, the day that the order was placed, and other details of the job specification. We also have the requested rating of the job as well as the grade the job received from the bank of reviewers. Table 1 shows the number of jobs requested at each quality rating and the quality grade of the work that the writer delivered to the customer. By exploiting variation in the ratings, we can examine the impact of the quality level delivered (assessed objectively by the reviewer), relative to the level that the customer requested, on customers' future transactional activity.

All observed transactions for a particular customer occur from the day of a customer's first order (t_1) until the end of our observation period (T). Treating the time of the initial trial as the beginning of the customer relationship is consistent with prior research in customer base analysis (Fader and Hardie 2001; Fader, Hardie, and Lee 2005a; Schmittlein, Morrison, and Colombo 1987). If a customer places x orders during that observation period, his or her observed "frequency" is equal to $x/(T - t_1)$. A customer's "recency" is t_x, the week of the most recently observed order. Each day is represented as one-seventh of a week.

Table 1
NUMBER OF JOBS REQUESTED AT EACH QUALITY RATING AND DELIVERED QUALITY GRADE

		Post Hoc Quality Grade				
		A/B	*C*	*D*	*E*	*Total*
Requested Rating	A	773	9	0	0	782
	B	8,784	614	6	1	9,405
	C	2,050	5,714	257	16	8,037
	D	1,668	3,270	814	83	5,835
	Total	13,275	9,607	1,077	100	24,059

To control for the possibility that some of the firm's earlier adopters might behave differently than customers whose first order came later, we divide the customer base into four cohorts on the basis of the week of the first order. The 588 customers who placed their first order during the first 33 weeks of our data are considered to be in the first cohort. The 568 customers who placed their first order between weeks 33 and 66 are assigned to the second cohort. The 911 customers who placed their first order between weeks 66 and 99 are assigned to the third cohort, and the 981 customers who placed their first order between weeks 99 and 130 are assigned to the fourth cohort.

Model Estimation

In this example, the transaction attributes represent the requested and delivered quality of the jobs. Although there are many functional forms that we could choose, we consider models of the form $\log q_k = \beta' z_k$, where z_k is a vector of job-specific covariates and β is a vector of coefficients. Effects that increase q_k increase the probability of churn.

The elements of z_k include the following:

- z_{first}: An indicator of the customer's first job;
- z_{coh2}, z_{coh3}, z_{coh4}: Indicators for time-invariant cohort effects;
- z_{Ak}, z_{Ck}, z_{Dk}: Indicators for the requested quality level of job k;
- z_{Lk}, z_{Hk}: Indicators for whether job k was lower (L) or higher (H) than the requested service level; and
- z_{LD}, z_{HD}, z_{LC}: Interactions among requested and delivered quality ratings.

All coefficients, except those on z_{Lk} and z_{Hk}, are stationary. The coefficients β_{Lk} and β_{Hk} represent the effect on the churn probability from missing or exceeding the specifications of job k. These effects can change from job to job according to the customer's recent experience. To capture how these sensitivities change, we define a set of six η parameters that affect β_{Lk} and β_{Hk} in the following ways:

(8) $$\beta_{L,k+1} = \beta_{Lk} + \eta_L + \eta_{LL} z_{Lk} + \eta_{HL} z_{Hk}, \text{ and}$$

(9) $$\beta_{H,k+1} = \beta_{Hk} + \eta_H + \eta_{LH} z_{Lk} + \eta_{HH} z_{Hk}.$$

The coefficient β_{Lk} changes by η_L regardless of the rating given to job k, capturing a drift in customers' sensitivity to receiving a lower-than-requested service level. The terms η_{LL}

and η_{HL} capture the extent to which customers' sensitivity to receiving a lower-than-requested service level on job k + 1 is affected by having received a lower (η_{LL}) or higher (η_{HL}) level of service than was requested for job k. The coefficient β_{Hk} evolves in a similar manner. By allowing β to be dynamic in this way, we allow customers' experiences to influence their responses to service quality (Bolton 1998). For example, if η_{LL} were positive, then after having a bad experience with the firm, a customer would be even more sensitive to subsequent bad experiences.

To assess the role of service quality on churn propensities, we tested three variants of the model. Model 3 is the full model as described. Model 1 is a "baseline" model that ignores all service quality effects. Model 2 is similar to Model 3, with all of the insignificant η parameters removed.

Table 2 contains descriptions of the model parameters, along with maximum likelihood estimates and standard errors. The subscripts for the elements of β in the table correspond to those of z. The most noteworthy estimates are those on β_{L1} and η_{LL}, which are both positive. This result suggests that, as we expected, missing the requested level of service for the first job increases the probability of churn. It also suggests that the magnitude of that effect will be larger for the next job with a missed service level. Thus, the churn probabilities increase across jobs when customers repeatedly receive lower-than-requested service.

We also observe that β_{H1}, and the associated η parameters, are not significantly different from zero. This asymmetry in the effect of service quality on customers' tendency to churn is consistent with losses looming larger than gains (Hardie, Johnson, and Fader 1993; Kahneman and Tversky 1979). Our findings are also in line with prior research by Bolton (1998), who finds that perceived losses adversely affect the duration of a customer's relationship in a contractual setting, whereas perceived gains do not have a significant impact on the duration of the relationship.

Model Assessment

We compare the performance of the three model specifications using a series of assessments. A likelihood ratio test suggests a weak preference for Model 2 over Model 1 ($\chi^2_6 = 11.0, p = .088$); we cannot infer a preference for Model 3 over Model 2 ($\chi^2_5 = 1.29, p = .936$) or for Model 3 over Model 1 ($\chi^2_9 = 10.1, p = .342$). One shortcoming of relying only on the likelihood ratio test, however, is that it does not consider the extent to which the incorporation of transaction attributes improves forecasting performance. Thus, in addition to the likelihood ratio test, we compare model performance using two additional measures of fit.

We calculated the mean absolute percentage error (MAPE) associated with the models' prediction of the weekly number of repeat orders made by customers in our sample and the root mean square error (RMSE) for the models' predictions of the distribution of the number of orders made by the sample. We find that the MAPE is similar across model specifications, with Models 2 and 3 having slightly lower errors than Model 1. In terms of the RMSE, we again find evidence to suggest that the transaction attributes contribute to model fit. Models 2 and 3 have lower RMSEs compared with Model 1 for the data used to estimate the model during the calibration and forecasting periods. Using a holdout sample for cross-validation reveals that although Model 2 has a lower RMSE than Model 1, Model 3 has a higher RMSE during the forecasting period, which would suggest that the model is overparameterized.

Details of these posterior predictive tests appear in the Web Appendix. We believe that the posterior predictive tests, taken together with the likelihood ratio test, provide

Table 2
PARAMETER ESTIMATES

	Model 1		*Model 2*		*Model 3*		
	Est.	*SE*	*Est.*	*SE*	*Est.*	*SE*	*Description*
r	.90	.04	.90	.04	.90	.04	Shape parameter on λ
a	.77	.04	.77	.04	.77	.04	Scale parameter on λ
s	1.09	.07	1.08	.07	1.10	.07	Shape parameter on θ
b	1.13	.19	1.14	.19	1.17	.20	Scale parameter on θ
β_{first}	−.22	.07	−.22	.07	−.21	.07	Effect of customer's first job
β_{coh2}	−1.15	.11	−1.16	.11	−1.15	.11	Fixed effect for cohort 2
β_{coh3}	−1.00	.10	−1.01	.11	−1.00	.11	Fixed effect for cohort 3
β_{coh4}	−.81	.11	−.81	.11	−.80	.11	Fixed effect for cohort 4
β_A	.88	.11	.89	.11	.89	.11	Effect for requested quality level A
β_C	−.30	.06	−.27	.07	−.27	.07	Effect for requested quality level C
β_D	−.51	.08	−.36	.15	−.37	.15	Effect for requested quality level D
β_{L1}			.19	.13	.21	.14	Effect for rating being "lower" than requested
β_{H1}			.01	.09	.01	.10	Effect for rating being "higher" than requested
β_{LD}			.21	.36	.24	.38	Interaction effect between "lower" and requested level D
β_{HD}			−.18	.17	−.15	.17	Interaction effect between "higher" and requested level D
β_{LC}			−.53	.26	−.52	.26	Interaction effect between "higher" and requested level C
η_L					−.01	.01	Evolution parameter on β_L
η_H					.00	.01	Evolution parameter on β_H
η_{LL}			.18	.12	.29	.21	Evolution parameter on β_L after a "lower" rating
η_{LH}					.05	.13	Evolution parameter on β_H after a "lower" rating
η_{HL}					.01	.04	Evolution parameter on β_L after a "higher" rating
η_{HH}					−.01	.02	Evolution parameter on β_H after a "higher" rating

evidence that transaction attributes improve model performance and contribute to forecasting accuracy. Based on these analyses, we focus on the results using Model 2 for the remainder of our discussion.

To provide a better sense for how well the proposed model captures customers' observed behavior, Figure 1 illustrates model fit at the aggregate level. Figure 1, Panel A, plots the cumulative and incremental number of weekly orders for both in-sample and holdout populations. The vertical lines at Week 130 divide the calibration and forecast time periods. We used only data to the left of the lines for estimating the model parameters and included only those customers whose initial order was before Week 130. Model 2 does well in tracking the number of orders from week to week. In Figure 1, Panel B, we compare the histogram of per-customer order counts with the distribution of counts from Model 2. Again, Model 2 seems to fit the data rather well.

At the customer level, one managerially relevant test statistic is the probability that a customer will place an order sometime in the future. Although the probability of being active, $P(\mathcal{A})$, is a commonly used construct in customer base analysis, we cannot use it as a model-checking tool because we cannot observe the customer's activity state directly. Instead, the appropriate metric is $P^*(0) = P(X^*(t^*) = 0|x, t_x, \cdot)$, the posterior probability that a customer will place no orders during a forecast period. We test how well Model 2 predicts which customers will order during the forecast period using a calibration plot. First, we assign each customer to one of 15 "bins" according to the customer's posterior $P^*(0)$. A customer is assigned to bin i if $(i - 1)/15 < P^*(0) \le i/15$ for i = 1, ..., 15. Next, we compute the observed proportion of customers in each bin who do not place an order during the forecast period. We consider the model to be well-calibrated if the predicted probabilities and observed proportions are aligned. Figure 2 confirms that they are. Each dot represents the membership of the bin. The x-coordinate is the midpoint of the bin, and the y-coordinate is the observed incidence of "no orders" for the members of that bin. "Perfect" calibration would have occurred if all of the dots fell exactly on the 45-degree line. Of course, we expect some random variation around this line, so we can still be confident that Model 2 forecasts the incidence of future orders at the customer level quite well.

Forecasting Quality Data

In the "Conditional Expectations and DERT" subsection, we discussed the need to model the sequences of covariates for the purpose of estimating conditional expectations and DERT rather than simply plugging in a fixed value. The specifics of such a model depend on the context. For this data set, there are two sources of variation in the covariates: the service level that a customer requested and the level that was delivered. For the "requested" model, we assume that each customer has a latent, stationary probability of placing A-, B-, C-, or D-rated orders. This probability vector is heterogeneous and varies across customers according to a Dirichlet distribution, allowing some customers to always choose the same rating while other customers can vary more in their choices. Under the further assumption of a zero-order choice process, we can infer the likelihood of a particular order pattern by using the maximum likelihood estimates of a Dirichlet multinomial mixture model.

The Dirichlet multinomial parameters for this data set are .04, .26, .22, and .12 for ordered ratings A, B, C, and D, respectively. These low values (all less than 1) suggest high polarization in the Dirichlet mixing distribution; even though there may be variation across customers, a single customer is likely to place orders for the same level of quality across the jobs (s)he orders. To simulate hypothetical orders, we sample a probability vector from each customer's posterior Dirichlet distribution. Because most customers order at the same rating level every time, these posterior probabilities are even more concentrated on a single quality level compared with the choice probabilities of the prior distribution. We then use the empirical distribution of the job grades for each rating level to get the service level delivered for each simulated job. In general, we find that there is little variation in B_k across simulated sequences of z_k.

HOW TRANSACTIONAL PATTERNS AFFECT DERT AND INCREMENTAL DERT

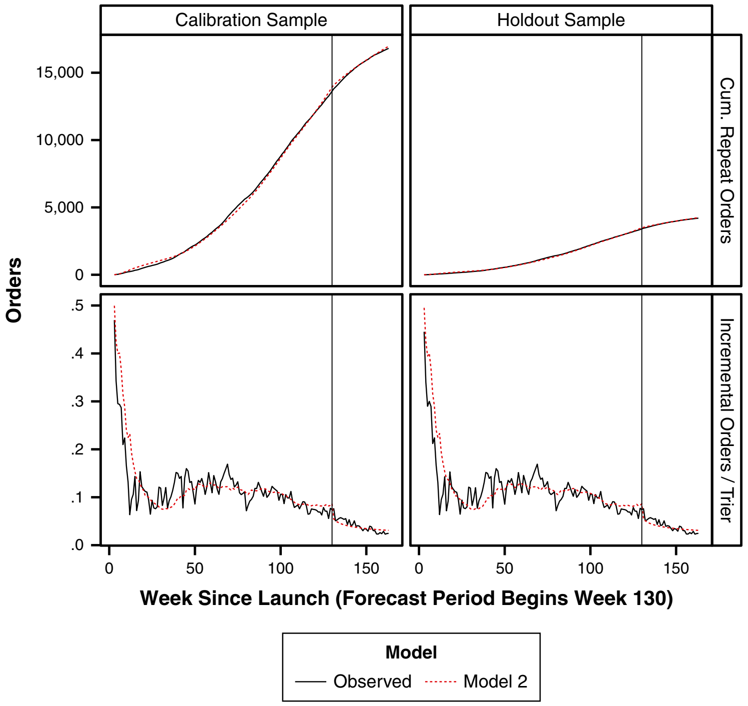
Using the parameter estimates of Model 2, we can examine how information from transaction attributes affects expectations of the future transactional activity of heterogeneous customers. Specifically, we examine how falling short of the expected level of service affects DERT for customers with different frequency (x) and recency (t_x) data. Figure 3 plots the contours that connect the same levels of DERT at T = 130 for hypothetical customers who placed the first order at time $t_1 = 1$ and who requested quality grades B, C, or D for the most recent order. For each of requested service levels, we consider delivered service levels that are lower than, or the same as, what was requested for that most recent order. We assume that the level of service delivered was the same as what the customer requested for all other orders. These isovalue curves are similar in spirit to those introduced by Fader, Hardie, and Lee (2005b) for the Pareto/NBD model. Because DERT is a posterior expectation based on the likelihood that a customer remains active, as anticipated, we observe "backward-bending" contours (Fader, Hardie, and Lee 2005b). When the number of orders is large and the most recent order was in the distant past, it is more likely that the customer has already become inactive than it would be if the x^{th} order were made more recently.

The iso-DERT curves in Figure 3 reveal the relationship between customers' transaction histories and expected future activity, but they do not show the *incremental* effect of missing the requested service level. To assess the incremental impact of deviations in the level of service on expected future transactions, we calculate the difference in DERT between what we would expect when the most recent job is rated at the same level of service the customer ordered and what we would expect if the level of service delivered was lower. We refer to this difference as "incremental DERT." This metric offers a long-term assessment of the marketing investment's impact and forms an upper bound on the amount the firm should invest. In transactions for which incremental DERT is small, it may not be worth the firm's effort to monitor and evaluate the service encounters. Transactions for which the incremental DERT is large, however, may warrant additional resources to ensure that the appropriate level of service is delivered. In our empirical context, incremental DERT can be viewed as the penalty

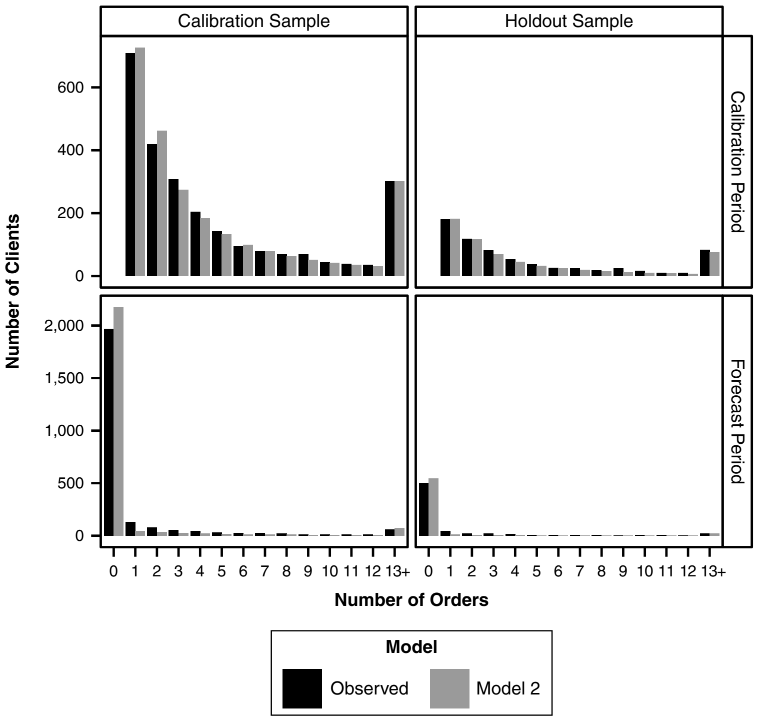
Figure 1

FIT AND FORECAST ASSESSMENT FOR MODEL 2

A: Weekly Incremental Repeat Orders per Previously Acquired Customer[a]



B: Observed and Predicted Histograms of Orders

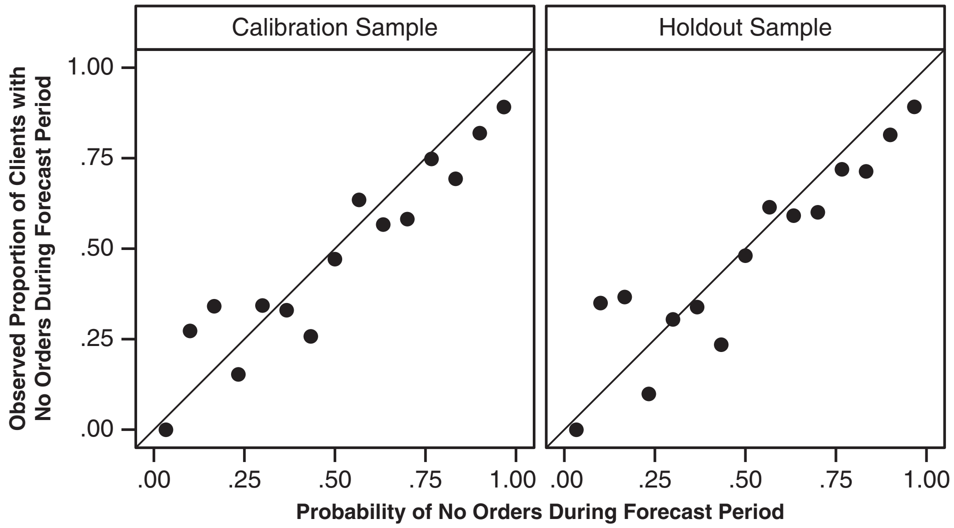


[a] The vertical line is at Week 130, the end of the calibration period and the start of the forecast period.

Notes: Cum. = cumulative.

Figure 2

PREDICTED VERSUS OBSERVED PROBABILITY OF A PARTICULAR CUSTOMER MAKING NO ORDERS DURING THE FORECAST PERIOD



associated with delivering a level of service lower than what a customer requested.

The expected cost of missing a requested service level for a brand new customer is the incremental DERT when $x = 1$ and $t_1 = t_x = T$. That is, immediately after the first job is ordered and a new customer comes under observation, incremental DERT is the number of discounted future transactions that we would expect to be lost if the delivered service level were below what was requested. For this particular data set, the percentage change is 2.9% for new customers who ordered a B-level job, 2.5% for a C-level job, and 2.2% for a D-level job. For existing customers, incremental DERT depends crucially on how many orders the customer has placed and how long ago the last order was made. Figure 4 shows this relationship in the form of "iso-incremental DERT" curves, in both absolute and percentage terms, for $t_1 = 1$ and $T = 130$ for a customer who has requested a service level of B. The levels of these curves represent the incremental DERT, evaluated at varying levels of frequency and recency. The darker contours represent differences that are more negative, for which the mismatch in service levels has more of an effect.[4]

In absolute terms, at a given level of frequency, falling short of the requested service level seems to have the greatest effect when recency is neither too high nor too low. Take the case of a customer who has conducted a given number of transactions. If the last transaction was conducted recently, it is likely that the customer is still active. Regardless of whether the service delivered was at or below the level requested, the recency of the transaction leads us to estimate a high DERT for that customer. Alternatively, if the most recent order was placed at a time in the distant past, the customer is more likely to have already defected, regardless of the service quality level, so we estimate a low DERT. In both of these cases, the difference in DERT between the case in which the delivered service matches what was requested and the case in which it is below what was requested will be small compared with the incremental DERT for a customer with moderate recency. For those customers, there is more uncertainty about whether the customer is still active. Small differences in attrition probabilities will yield more substantial decreases in DERT when the requested service level is missed. We observe that the range of incremental DERT is the widest at low or moderate order frequencies.

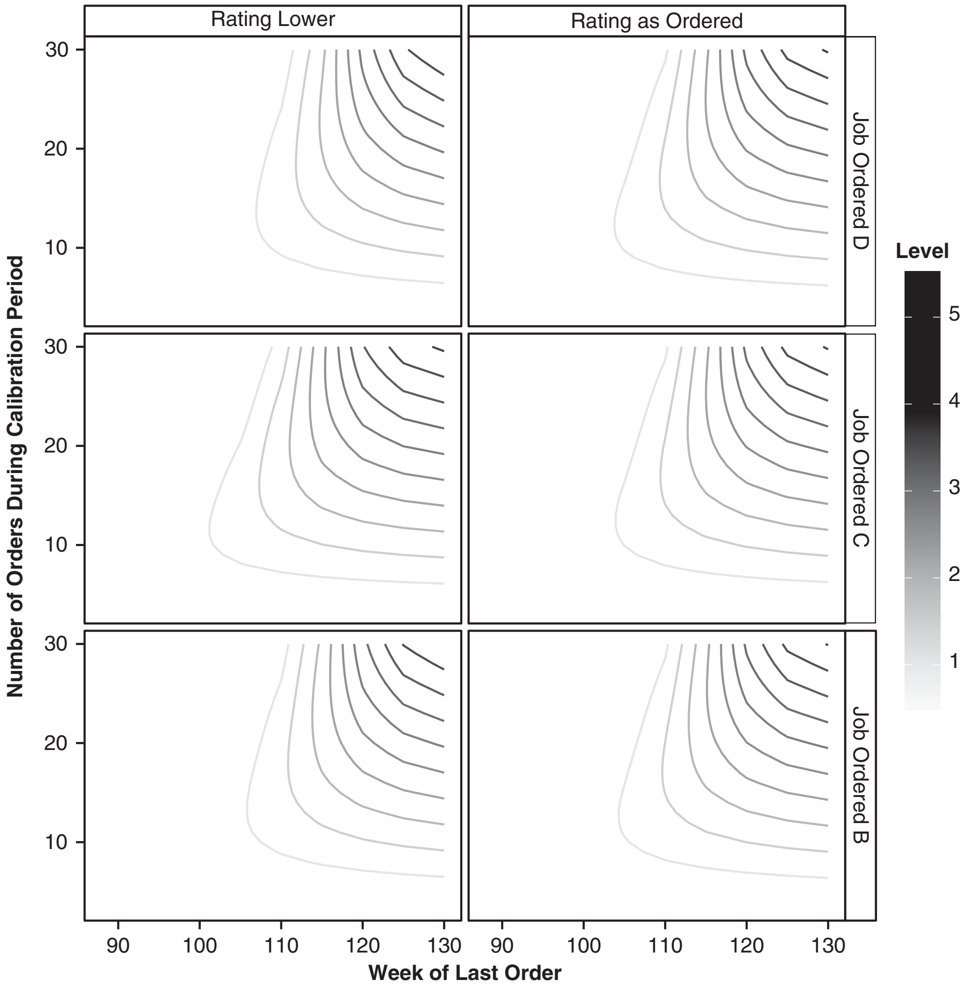
Figure 4, Panel B, illustrates the DERT contours in percentage terms. As Figure 4, Panel A, shows, for a given number of orders x, there are two values t_x that correspond to the same absolute difference in DERT. As DERT increases with recency for a given value of x (as Figure 3 illustrates), incremental DERT on a percentage basis will be smaller at higher values of recency.

Figure 4 demonstrates that the expected difference in a customer's remaining value between meeting and falling short of the requested service level depends on both the recency and frequency of the customer's transactions. When the goal of customer relationship management is to identify those customers for whom targeted marketing actions generate the greatest financial impact, our findings should caution against scoring customers solely on popular measures such as P(Alive) or customer lifetime value. Although the most recent and frequently transacting customers are likely to be the most valuable, the effect of transaction attributes is not much different from those of customers who last purchased a long time ago.

To provide another perspective on customers' remaining value, Figure 5 shows incremental DERT (absolute values in Figure 5, Panel A, and percentages in Figure 5, Panel B) as a function of frequency for select values of recency. That is, each curve corresponds to a vertical slice of either Figure 4, Panel A, or Figure 4, Panel B. For transaction

[4]We observe a similar pattern as depicted in Figure 4 for jobs that exceed the order specifications, except with positive differences in DERT. However, for this data set, the magnitude of the effect is very small.

Figure 3
DERT ISOVALUE CURVES FOR HYPOTHETICAL CUSTOMERS WHOSE FIRST ORDER WAS AT $t_1 = 1$



Notes: Both the left- and right-hand panels condition on the ordered job quality and whether the last job was rated lower than or the same as what the customer ordered. The x-axis of each panel is the week number of the most recent order (T = 130), and the y-axis is the number of observed orders.

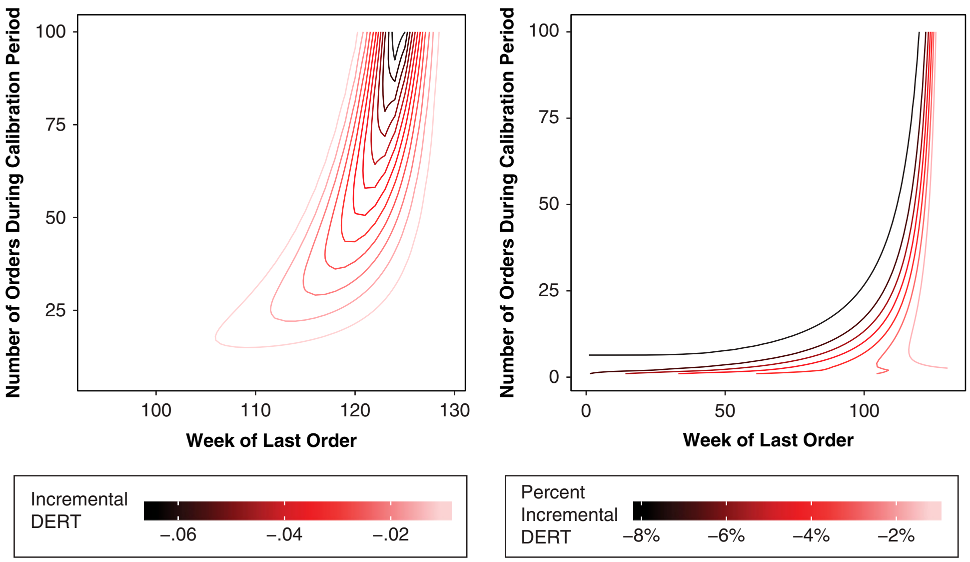
histories with lower recency (e.g., $t_x = 85$ and $t_x = 100$), we observe an initial increase in the magnitude of the incremental DERT (in absolute terms) as the number of orders increases. However, with a very large number of orders, the magnitude of incremental DERT decreases and approaches zero. This pattern reveals the interplay of two factors. First, the large number of orders is indicative of a high transaction rate (λ). For these customers, missing the requested service level puts a lucrative revenue stream at risk. Yet with a large number of orders, the low recency of the last order suggests that the customer may have already become inactive. Given the increased likelihood that the customer has already lapsed, the increased cost associated with missing the requested service level is lower, resulting in the reduced magnitude of incremental DERT. At the highest level of recency ($t_x = 130$), the customer is known to be still active, and DERT increases with the number of orders. We thus observe that the magnitude of incremental DERT increases with frequency, but at a much slower rate.

Figure 5, Panel B, presents a similar analysis in percentage terms. At the highest level of recency ($t_x = 130$), when the customer has just conducted a transaction and is known to still be active, DERT increases with frequency. Therefore, although incremental DERT is relatively flat in absolute terms, it decreases when expressed as a percentage of DERT. For the remaining three levels of recency, higher transaction frequency ultimately results in a higher magnitude of incremental DERT as a percentage. As the number of transactions increases along these three curves, so too does the likelihood that these customers have already become inactive, thereby driving down DERT. When incremental DERT (in absolute terms) is compared with a smaller base, it increases as a percentage.

Prior research has suggested that the value of customer relationship management tools lies in their ability to facilitate learning about customers (Mithas, Krishnan, and Fornell 2005) and that this value is limited on the basis of the quality of available information the firm has about its

Figure 4
ISO-INCREMENTAL DERT CURVES WHEN MISSING THE REQUESTED SERVICE LEVEL

A: Absolute Difference *B: Percentage Difference*

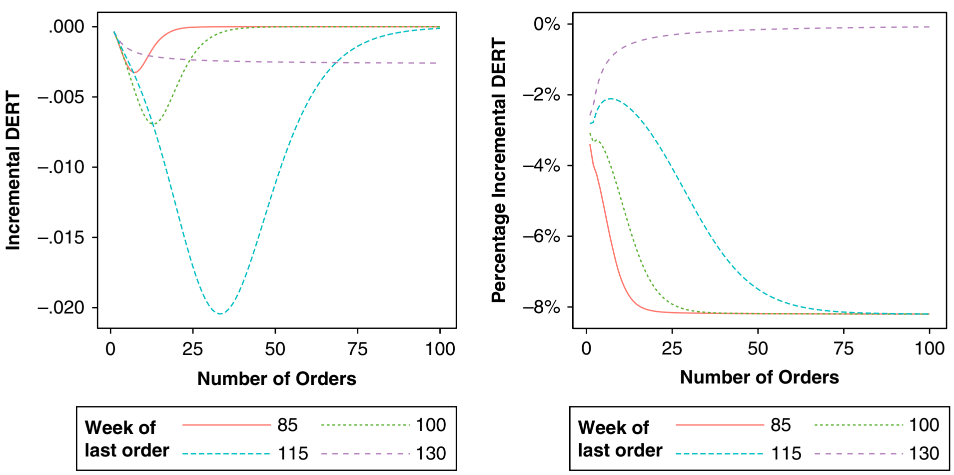


Notes: Observed recency is on the x-axis and frequency is on the y-axis. Each curve connects points for which the incremental DERT is the same.

customers (Homburg, Droll, and Totzek 2008). Our findings reveal that differences in customer value that are associated with variation in transaction attributes (e.g., the level of service delivered on a transaction) depend on customers' past transaction activity. If a customer is more obviously active or inactive, deviations from the requested service level provide little information on the customer's DERT. It is when there is increased uncertainty as to whether a customer is active or inactive that the level of service delivered relative to the level requested enables us to learn more about a customer's DERT. To the best of our knowledge, our research is the first to explore the value of transaction attributes using the latent attrition models frequently employed in customer base analysis and

Figure 5
ABSOLUTE AND PERCENTAGE INCREMENTAL DERT AS A FUNCTION OF FREQUENCY FOR DIFFERENT LEVELS OF RECENCY

A: Absolute Difference *B: Percentage Difference*



customer valuation and how past activity affects the differences in customer value associated with transaction attributes.

DISCUSSION

We present a flexible latent attrition model that incorporates transaction attributes in customer base analysis and describe an empirical application that defines those attributes as indicators for relative service quality. The model allows us to derive "incremental DERT," a metric of the discounted expected return on changes in those attributes. For transactions that are about to happen immediately, incremental DERT can serve as an upper bound on the amount a firm should invest to change one of those attributes. We also describe patterns in the relationship between incremental DERT and the customer's recency/frequency profile. Understanding these patterns and quantifying the effects enable managers to more accurately estimate DERT for all members of the customer base. Model parameters can be estimated using standard maximum likelihood methods, and DERT can be computed as a truncated summation.

Because firms use customer base analysis and customer valuation models to score and rank customers according to the value they hold for a firm (Wuebben and Von Wangenheim 2008), managers should be interested in tracking and compiling transaction attributes that may be informative of customers' future value. Acquiring such information, however, often entails a cost. For example, firms may incur costs to monitor their sales force to ensure compliance with required activities (John and Weitz 1989). Customers' purchase frequencies could be a simple heuristic to identify transactions that firms should be willing to invest more in monitoring. After all, such customers are likely to be among the most valuable customers to a firm. Our results, however, suggest that for such customers, the difference in discounted long-term value to the firm between meeting and falling short of the requested service level (reflected by incremental DERT) is low. For customers who have not conducted as many transactions, our estimates of their value to the firm are more sensitive to information on the level of service delivered. It is for these customers that estimates of long-term value are most subject to change if the level of service in a transaction falls short of specifications.

In our empirical context, we report that the magnitude of the effect of missing specified levels of service is larger than that of exceeding specified levels. This is consistent with the idea of losses looming larger than gains. Although the possibility of raising customers' future expectations has led extant research to question the wisdom of trying to delight customers by exceeding their expectations (Rust and Oliver 2000), our analysis suggests that falling short of the service customers have requested poses a greater risk to a customer's continued relationship with the firm and the firm's ability to capture the corresponding revenue stream.

Our framework provides a foundation for several promising directions of further research. Although we focus our attention on customer retention and the value of existing customers, a similar modeling approach could be employed to jointly investigate customer acquisition and retention (Musalem and Joshi 2009; Schweidel, Fader, and Bradlow 2008). Doing so could provide firms with guidance for how to balance marketing expenditures across the two activities (Reinartz, Thomas, and Kumar 2005). Another area to explore is the effect of strategic investments in service quality, a practice that the firm that supplied our data did not employ. Although it may be costly for a firm to invest in improving service encounters and monitoring these encounters for all customers, it may have resources to focus on select customers and could use incremental DERT as a criterion for selecting those target customers. If the costs of delivering better-than-expected service experiences are the same across customers, such an allocation rule would be equivalent to the firm putting its money where it will deliver the most "bang for the buck." These actions would be consistent with the management principle of "return on quality" (Rust, Zahorik, and Keiningham 1995). In many cases, we would need to account for the firm's actions when estimating the effect on retention probabilities (Manchanda, Rossi, and Chintagunta 2004; Schweidel and Knox 2013). To alleviate such concerns, future researchers might choose to proceed in this area with a carefully designed field experiment.

Another area for further research would be to develop additional means of accounting for variation in the effect of customer–firm touch points on future customer behavior. Although we rely on observable transaction attributes (i.e., the requested and evaluated service levels) as a means of evaluating the level of service, future work could explore whether such a measure could be inferred with additional information about the transaction. For example, in our empirical context, if the transaction data indicated the writer who completed each job, the firm might be able to evaluate the "quality" of each writer on the basis of the effect (s)he had on customer churn. Another example of a firm that offers a marketplace is eBay, which connects buyers and sellers. Evaluating sellers on the basis of the churn of buyers with whom they have recently interacted may provide an indication of which sellers should be rewarded versus which sellers are potentially costing the firm business.

The cost of conducting such an analysis is in the detail of the data that must be collected. Although, as in our analysis, the early latent attrition models in the marketing literature relied on recency and frequency as sufficient statistics, recognizing that the variation in customer tendencies that exists across transactions requires that data be tracked at the transaction level. It is an empirical question as to the extent to which incorporating such sources of variation into the analysis will affect managerial decisions. Because the answer to this question may vary from context to context, additional research is warranted that recognizes the costs associated with acquiring and analyzing data on customer–firm interactions across a range of empirical applications.

APPENDIX: DERIVATIONS

We use the definitions and symbols that are defined in Table A1. To derive the marginal likelihood in Equation

Table A1
DEFINITIONS OF SYMBOLS AND FUNCTIONS USED IN THE ARTICLE

$\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt$	Gamma function
$\gamma(k,\lambda) = \int_0^\lambda t^{k-1}e^{-t}dt$	Lower incomplete gamma function
$\mathcal{G}_z(r,a) = \gamma(r,az)/\Gamma(r)$	CDF of a gamma distribution with shape r and rate a
$d\mathcal{G}_z(r,a) = \frac{a^r}{\Gamma(r)}z^{r-1}e^{-az}$	Density of a gamma distribution with shape r and rate a
$\mathbb{B}(k,r) = \int_0^1 u^{k-1}(1-u)^{r-1}du$	Beta function
$\mathbb{B}(z;k,r) = \int_0^z u^{k-1}(1-u)^{r-1}du$	Incomplete beta function
$\tilde{\mathbb{B}}(z;k,r) = \mathbb{B}(z;k,r)/\mathbb{B}(k,r)$	Regularized incomplete beta function (equivalent to CDF of beta distribution with parameters k and r, evaluated at z)
${}_2F_1(a,b,c;z)$	Gaussian hypergeometric function
$P(\mathcal{A}\|\lambda,\theta)$, $P(\mathcal{A})$	Conditional and marginal probabilities that a customer has not yet churned by time T

3, we integrate the individual-level data likelihood in Equation 1 over two gamma densities, one for λ and one for θ.

(A1)
$$\begin{aligned}\mathcal{L} &= \int_0^\infty\int_0^\infty f(x,t_{2:x}|\lambda,\theta,q_{1:x})d\mathcal{G}_\lambda(r,a)d\mathcal{G}_\theta(s,b)\\
&= \int_0^\infty\int_0^\infty \lambda^{x-1}e^{-\lambda(t_x-t_1)-\theta B_{x-1}}\left[1-e^{-\theta q_x}\left(1-e^{-\lambda(T-t_x)}\right)\right]\\
&\quad\times\frac{a^r}{\Gamma(r)}\lambda^{r-1}e^{-a\lambda}\frac{b^s}{\Gamma(s)}\theta^{s-1}e^{-b\theta}d\lambda d\theta\\
&= \frac{a^r}{\Gamma(r)}\frac{b^s}{\Gamma(s)}\int_0^\infty\lambda^{r+x-2}e^{-\lambda(a+t_x-t_1)}d\lambda\int_0^\infty\theta^{s-1}e^{-\theta(b+B_{x-1})}d\theta\\
&\quad-\frac{a^r}{\Gamma(r)}\frac{b^s}{\Gamma(s)}\int_0^\infty\lambda^{r+x-2}e^{-\lambda(a+t_x-t_1)}d\lambda\int_0^\infty\theta^{s-1}e^{-\theta(b+B_x)}d\theta\\
&\quad+\frac{a^r}{\Gamma(r)}\frac{b^s}{\Gamma(s)}\int_0^\infty\lambda^{r+x-2}e^{-\lambda(a+T-t_1)}d\lambda\int_0^\infty\theta^{s-1}e^{-\theta(b+B_x)}d\theta\\
&= \frac{\Gamma(r+x-1)}{\Gamma(r)}\frac{a^r}{(a+t_x-t_1)^{r+x-1}}\left(\frac{b}{b+B_{x-1}}\right)^s\\
&\quad\times\left\{1-\left(\frac{b+B_{x-1}}{b+B_x}\right)^s\left[1-\left(\frac{a+t_x-t_1}{a+T-t_1}\right)^{r+x-1}\right]\right\}.\end{aligned}$$

By rearranging terms, we can write the marginal likelihood equivalently as

$$\text{(A2)}\quad \begin{aligned}\mathcal{L} &= \frac{\Gamma(r+x-1)}{\Gamma(r)}\frac{a^r}{(a+T-t_1)^{r+x-1}}\left(\frac{b}{b+B_x}\right)^s\\
&\quad\times\left\{1-\left(\frac{a+T-t_1}{a+t_x-t_1}\right)^{r+x-1}\left[1-\left(\frac{b+B_x}{b+B_{x-1}}\right)^s\right]\right\}.\end{aligned}$$

We can also compute the expected number of orders for any randomly chosen customer in the population. Our approach draws inspiration from Fader, Hardie, and Lee (2005a, Section 4.3). Let τ be the time of the job immediately after which the customer churns. Therefore, conditional on k, the probability that τ is sometime after t is equal to the probability of surviving all k transactions that occurred before t. This survival probability is $e^{-\theta B_k}$. The probability of making k transactions is a shifted Poisson (because k starts at 1). By summing all possible values of k, we get

$$\text{(A3)}\quad P(\tau>t) = \sum_{k=1}^\infty\frac{(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}e^{-\theta B_k}.$$

By differentiating Equation A3 with respect to t, we get the probability density function of τ.

$$\text{(A4)}\quad \begin{aligned}g(\tau) &= \sum_{k=1}^\infty\frac{e^{-\theta B_k}\lambda^{k-1}}{(k-1)!}\frac{d}{dt}\left[t^{k-1}e^{-\lambda t}\right]\\
&= e^{-\lambda\tau}\sum_{k=1}^\infty\frac{e^{-\theta B_k}(\lambda\tau)^{k-1}}{(k-1)!}\left[\lambda-\frac{k-1}{\tau}\right].\end{aligned}$$

To calculate the expected number of transactions for an individual customer, we have to consider two cases. In the first case, the customer survives until time t, so the expected number of repeat orders (not including the first order) is λt multiplied by the probability that τ > t. In the second case, the customer defects at time τ, which is sometime before t. In this case, the expected number of repeat transactions is λτ. Because τ is unknown, we can calculate the expected number of repeat transactions by integrating τ over the entire interval in question, with respect to g(τ).

(A5)
$$\begin{aligned}E[X(t)|\lambda,\theta] &= \lambda t\left[\sum_{k=1}^\infty\frac{(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}e^{-\theta B_k}\right]\\
&\quad+\lambda\int_0^t\tau e^{-\lambda\tau}\sum_{k=1}^\infty\frac{e^{-\theta B_k}(\lambda\tau)^{k-1}}{(k-1)!}\left[\lambda-\frac{k-1}{\tau}\right]d\tau\\
&= \lambda t\left[\sum_{k=1}^\infty\frac{(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}e^{-\theta B_k}\right]\\
&\quad+\sum_{k=1}^\infty\frac{\lambda^k e^{-\theta B_k}}{(k-1)!}\int_0^t e^{-\lambda\tau}\tau^k\left[\lambda-\frac{k-1}{\tau}\right]d\tau\\
&= \left[\sum_{k=1}^\infty\frac{(\lambda t)^k e^{-\lambda t}}{(k-1)!}e^{-\theta B_k}\right]\\
&\quad+\sum_{k=1}^\infty\frac{e^{-\theta B_k}}{(k-1)!}\left[\gamma(k,\lambda t)-(\lambda t)^k e^{-\lambda t}\right]\\
&= \sum_{k=1}^\infty\frac{\gamma(k,\lambda t)}{\Gamma(k)}e^{-\theta B_k}.\end{aligned}$$

To get the prior expectation for a randomly chosen member of the population (Equation 4), we integrate Equation A5 over the prior densities of λ and θ.

(A6)

$$E[X(t)] = \sum_{k=1}^{\infty} \frac{a^r}{\Gamma(r)} \frac{b^s}{\Gamma(s)} \frac{1}{\Gamma(k)} \int_0^{\infty} \gamma(k, \lambda t)\lambda^{r-1} e^{-a\lambda} \times \int_0^{\infty} \theta^{s-1} e^{-\theta(B_k + b)} d\theta\, d\lambda$$

$$= \sum_{k=1}^{\infty} \frac{a^r}{\Gamma(r)} \frac{1}{\Gamma(k)} \left(\frac{b}{b + B_k}\right)^s \int_0^{\infty} \gamma(k, \lambda t)\lambda^{r-1} e^{-\alpha\lambda} d\lambda.$$

To solve the last integral in Equation A6, we apply Equation 6.455.1 in Gradshteyn and Ryzhik (2000), which expresses the integral in terms of a Gaussian hypergeometric function.

(A7)
$$E[X(t)] = \sum_{k=1}^{\infty} \frac{a^r}{\Gamma(r)} \frac{1}{\Gamma(k)} \times \left(\frac{b}{b + B_k}\right)^s \frac{t^k \Gamma(r + k)}{k(a + t)^{r+k}} \; {}_2F_1\left(1, r + k; k + 1; \frac{t}{a + t}\right).$$

We can simplify this expression using identities in two sections of the *NIST Handbook of Mathematical Functions* (Olver et al. 2010). First, we transform the hypergeometric function using the identity in Section 15.8.1. Then, we apply the hypergeometric representation of an incomplete beta function from Section 8.17.9. Finally, we regularize the incomplete beta function to get Equation 4.

(A8)
$$E[X(t)] = \sum_{k=1}^{\infty} \frac{a^r}{\Gamma(r)} \frac{1}{\Gamma(k)} \left(\frac{b}{b + B_k}\right)^s \frac{t^k \Gamma(r + k)}{k(a + t)^{r+k}} \times \left(\frac{a + t}{a}\right)^r \; {}_2F_1\left(k, 1 - r; k + 1; \frac{t}{a + t}\right)$$

$$= \sum_{k=1}^{\infty} \frac{\Gamma(r + k)}{\Gamma(r)\Gamma(k)} \left(\frac{b}{b + B_k}\right)^s \mathbb{B}\left(\frac{t}{a + t}; k, r\right)$$

$$= \sum_{k=1}^{\infty} \left(\frac{b}{b + B_k}\right)^s \tilde{\mathbb{B}}\left(\frac{t}{a + t}; k, r\right).$$

Note that the regularized incomplete beta function is equivalent to the CDF of a beta distribution.

Given a customer's transaction history, we can derive the joint posterior density of λ and θ by applying Bayes' theorem.

(A9)

$$g(\lambda, \theta | x, t_1 \ldots t_x) = \frac{1}{\mathcal{L}} f(x, t_{2:x} | \lambda, \theta) d\mathcal{G}_{\lambda}(r, a) d\mathcal{G}_{\theta}(s, b)$$

$$= \frac{1}{\mathcal{L}} \frac{a^r b^s}{\Gamma(r)\Gamma(s)} \lambda^{r+x-2} e^{-\lambda(a+t_x-t_1)} \theta^{s-1} e^{-\theta(b+B_{x-1})} \times \left[1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)\right]$$

$$= d\mathcal{G}_{\lambda}(r + x - 1, a + t_x - t_1) d\mathcal{G}_{\theta}(s, b + B_{x-1}) \times \frac{1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)}{1 - \left(\frac{b + B_{x-1}}{b + B_x}\right)^s \left[1 - \left(\frac{a + t_x - t_1}{a + T - t_1}\right)^{r+x-1}\right]}.$$

An important quantity of interest is the probability that a customer is still "alive" at the end of the observation period. At time T, the customer is in one of two possible states. One state is that the customer churned after the xth transaction. This occurs with probability p_x. The other state is that the customer survived the last transaction but has not purchased since. This occurs with probability $(1 - p_x)e^{-\lambda(T-t_x)}$. Therefore, the probability of the customer's being alive at time T, conditional on purchase history, is

(A10)

$$P(\mathcal{A}|\lambda, \theta) = \frac{(1 - p_x)e^{-\lambda(T-t_x)}}{p_x + (1 - p_x)e^{-\lambda(T-t_x)}} = \frac{e^{-\theta q_x - \lambda(T-t_x)}}{1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)}.$$

By integrating Equation A10 across the posterior density in Equation A9, we get Equation 5.

(A11)

$$P(\mathcal{A}) = \int_0^{\infty}\int_0^{\infty} \frac{e^{-\theta q_x - \lambda(T-t_x)}}{1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)} \frac{(a + t_x - t_1)^{r+x-1}}{\Gamma(r + x - 1)} \times \lambda^{r+x-2} e^{-\lambda(a+t_x-t_1)} \frac{(b + B_{x-1})^s}{\Gamma(s)} \theta^{s-1} e^{-\theta(b+B_{x-1})} \times \frac{1 - e^{-\theta q_x}\left(1 - e^{-\lambda(T-t_x)}\right)}{1 - \left(\frac{b + B_{x-1}}{b + B_x}\right)^s \left[1 - \left(\frac{a + t_x - t_1}{a + T - t_1}\right)^{r+x-1}\right]} d\lambda\, d\theta$$

$$= \frac{(a + t_x - t_1)^{r+x-1}}{\Gamma(r + x - 1)} \frac{(b + B_{x-1})^s}{\Gamma(s)} \times \left\{1 - \left(\frac{b + B_{x-1}}{b + B_x}\right)^s \left[1 - \left(\frac{a + t_x - t_1}{a + T - t_1}\right)^{r+x-1}\right]\right\}^{-1} \times \int_0^{\infty} \lambda^{r+x-2} e^{-\lambda(a + T - t_1)} d\lambda \int_0^{\infty} \theta^{s-1} e^{-\theta(b+B_x)} d\theta$$

$$= \left(\frac{a + t_x - t_1}{a + T - t_1}\right)^{r+x-1} \left(\frac{b + B_{x-1}}{b + B_x}\right)^s \times \left\{1 - \left(\frac{b + B_{x-1}}{b + B_x}\right)^s \left[1 - \left(\frac{a + t_x - t_1}{a + T - t_1}\right)^{r+x-1}\right]\right\}^{-1}$$

$$= \left\{1 - \left(\frac{a + T - t_1}{a + t_x - t_1}\right)^{r+x-1} \left[1 - \left(\frac{b + B_x}{b + B_{x-1}}\right)^s\right]\right\}^{-1}.$$

Through some straightforward manipulation of terms, we can also write $P(\mathcal{A})$ in terms of the marginal likelihood.

(A12)
$$P(\mathcal{A}) = \frac{1}{\mathcal{L}} \frac{\Gamma(r + x - 1)}{\Gamma(r)} \frac{a^r}{(a + T - t_1)^{r+x-1}} \left(\frac{b}{b + B_x}\right)^s.$$

Now we can compute the expected number of transactions for a specific customer, given an observed transaction history. Let $X(t^*)$ be the number of purchases in the next period of duration t^* (i.e., during the interval from T to $T + t^*$). Given a customer's observed history and individual-level parameters, the expected number of orders during the next t^* weeks is the probability that the customer is still "alive" at time T multiplied by the prior expectation in Equation A8.

(A13)
$$E\left[X(t^*)\middle| x, , t_x, \theta, \lambda\right] = \frac{e^{-\theta e^{\beta' z_x} - \lambda(T-t_x)}}{1 - e^{-\theta e^{\beta' z_x}}\left(1 - e^{-\lambda(T-t_x)}\right)} \sum_{k=1}^{\infty} \frac{\gamma(k, \lambda t^*)}{\Gamma(k)} e^{-\theta B_k}.$$

Equation 6 results from integrating λ and θ in Equation A13 over the posterior density in Equation A9.

(A14)

$$
\begin{aligned}
E[X(t^*)|x,t_x] &= \frac{(a+t_x-t_1)^{r+x-1}}{\Gamma(r+x-1)}\frac{(b+B_{x-1})^s}{\Gamma(s)} \\
&\times \left\{1-\left(\frac{b+B_{x-1}}{b+B_x}\right)^s\left[1-\left(\frac{a+t_x-t_1}{a+T-t_1}\right)^{r+x-1}\right]\right\}^{-1} \\
&\times \sum_{k=1}^{\infty}\int_0^{\infty}\frac{\gamma(k,\lambda t^*)}{\Gamma(k)}\lambda^{r+x-2}e^{-\lambda(a+T-t_1)}d\lambda \\
&\times \int_0^{\infty}\theta^{s-1}e^{-\theta(b+B_x-B_k)}\,d\theta \\
&= \left\{1-\left(\frac{b+B_{x-1}}{b+B_x}\right)^s\left[1-\left(\frac{a+t_x-t_1}{a+T-t_1}\right)^{r+x-1}\right]\right\}^{-1} \\
&\times \sum_{k=1}^{\infty}\frac{(a+t_x-t_1)^{r+x-1}}{\Gamma(r+x-1)\Gamma(k)}\left(\frac{B_x+b}{B_x+B_k+b}\right)^s \\
&\times \frac{t^{*k}\Gamma(r+x+k-1)}{k(t^*+a+T-t_1)^{r+x+k-1}} \\
&\times {}_2F_1\left(1,r+x+k-1;k+1;\frac{t^*}{t^*+a+T-t_1}\right) \\
&= P(\mathcal{A})\times\sum_{k=1}^{\infty}\left(\frac{B_x+b}{B_x+B_k+b}\right)^s\tilde{\mathbb{B}} \\
&\times \left(\frac{t^*}{t^*+a+T-t_1};k,r+x-1\right).
\end{aligned}
$$

For further details, see the Web Appendix.

REFERENCES

Abe, Makoto (2009), "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541–53.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45–65.

Boulding, William, Ajay Kalra, and Richard Staelin (1999), "The Quality Double Whammy," *Marketing Science*, 18 (4), 463–84.

———, ———, ———, and Valarie A. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions," *Journal of Marketing Research*, 30 (February), 7–27.

Braun, Michael and David A. Schweidel (2011), "Modeling Customer Lifetimes with Multiple Causes of Churn," *Marketing Science*, 30 (5), 881–902.

Fader, Peter S. and Bruce G.S. Hardie (2001), "Forecasting Repeat Sales at CDNOW: A Case Study," *Interfaces*, 31 (2), S94–S107.

———, ———, and Ka Lok Lee (2005a), "Counting Your Customers the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275–84.

——— ——— and ——— (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (November), 415–30.

———, ———, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, 29 (6), 1086–108.

Gradshteyn, I.S. and I.M. Ryzhik (2000), *Table of Integrals, Series, and Products*. San Diego: Academic Press.

Grassia, A. (1977), "On a Family of Distributions with Argument Between 0 and 1 Obtained by Transformation of the Gamma and Derived Compound Distributions," *Australian Journal of Statistics*, 19 (2), 108–14.

Griffiths, David and Christine Schafer (1981), "Closeness of Grassia's Transformed Gammas and the Beta Distribution," *Australian Journal of Statistics*, 23 (2), 240–46.

Hardie, Bruce G.S., Eric J. Johnson, and Peter S. Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12 (4), 378–94.

Ho, Teck-Hua, Young-Hoon Park, and Yong-Pin Zhou (2006), "Incorporating Satisfaction into Customer Value Analysis: Optimal Investment in Lifetime Value," *Marketing Science*, 25 (3), 260–77.

Homburg, Christian, Mathias Droll, and Dirk Totzek (2008), "Customer Prioritization: Does It Pay Off, and How Should It Be Implemented?" *Journal of Marketing*, 72 (September), 110–30.

IBM (2011), "From Stretched to Strengthened: Insights from the Global Chief Marketing Officer Study," executive summary, (accessed October 14, 2015), [available at http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=XB&htmlfid=GBE03433USEN].

John, George and Barton Weitz (1989), "Salesforce Compensation: An Empirical Investigation of Factors Related to Use of Salary Versus Incentive Compensation," *Journal of Marketing Research*, 26 (February), 1–14.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–91.

Knox, George and Rutger van Oest (2014), "Customer Complaints and Recovery Effectiveness: A Customer Base Approach," *Journal of Marketing*, 78 (September), 42–57.

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

Mithas, Sunil, M.S. Krishnan, and Claes Fornell (2005), "Why Do Customer Relationship Management Applications Affect Customer Satisfaction?" *Journal of Marketing*, 69 (October), 201–9.

Musalem, Andrés and Yogesh V. Joshi (2009), "How Much Should You Invest in Each Customer Relationship? A Competitive Strategic Approach," *Marketing Science*, 28 (3), 555–65.

Olver, Frank W.J., Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, eds. (2010), *NIST Handbook of Mathematical Functions*. New York: Cambridge University Press.

Pazgal, Amit and David Soberman (2008), "Behavior-Based Discrimination: Is It a Winning Play, and If So, When?" *Marketing Science*, 27 (6), 977–94.

Reinartz, Werner, 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., J. Jeffrey Inman, Jianmin Jia, and Anthony Zahorik (1999), "What You Don't Know About Customer-Perceived Quality: The Role of Customer Expectation Distributions," *Marketing Science*, 18 (1), 77–92.

——— and Richard L. Oliver (2000), "Should We Delight the Customer?" *Journal of the Academy of Marketing Science*, 28 (1), 86–94.

———, Anthony J. Zahorik, and Timothy Keiningham (1995), "Return on Quality (ROQ): Making Service Quality Financially Accountable," *Journal of Marketing*, 59 (April), 58–70.

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

Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "A Bivariate Timing Model of Customer Acquisition and Retention," *Marketing Science*, 27 (5), 829–43.

——— and George Knox (2013), "Incorporating Direct Marketing Activity into Latent Attrition Models," *Marketing Science*, 32 (3), 471–87.

Shin, Jiwoong and K. Sudhir (2010), "A Customer Management Dilemma: When Is It Profitable to Reward One's Own Customers?" *Marketing Science*, 29 (4), 671–89.

Villas-Boas, J. Miguel (1999), "Dynamic Competition with Customer Recognition," *RAND Journal of Economics*, 30 (4), 604–31.

Wuebben, Markus and Florian von Wangenheim (2008), "Instant Customer Base Analysis: Managerial Heuristics Often 'Get It Right'," *Journal of Marketing*, 72 (May), 82–93.