Journal of Service Research

http://jsr.sagepub.com

Targeting Customers with Statistical and Data-Mining Techniques

James H. Drew, D. R. Mani, Andrew L. Betz and Piew Datta

Journal of Service Research 2001; 3; 205
DOI: 10.1177/109467050133002

Targeting Customers With Statistical and Data-Mining Techniques

James H. Drew
D. R. Mani
Andrew L. Betz
Piew Datta
Verizon Laboratories Incorporated

Operationalizing a relationship management program requires a retention strategy that is sensitive to an individual customer's position in the service life cycle, while being financially sound for the provider. To this end, estimating a customer's hazard function and remaining tenure with the company can lead to important insights into marketing tactics and constitute fundamental building blocks for methods of targeting important customers. The authors describe a way of estimating these quantities using a combination of statistical and data-mining techniques. The resulting customer hazard information leads to a generalization of lifetime value (GLTV) that explicitly accounts for company actions and their success in relationship management.

One of the central tenets of relationship management is the development of marketing and retention strategies that are customized to customers' positions in their service usage life cycle (Aaker, Kumar, and Day 1998). In particular, it is an important operational tactic to target customers for special treatment based on their anticipated future value to the company. A basic concept for the implementation of these strategies is the estimation of some form of the value of a customer during that usage lifetime, and this has been frequently operationalized as lifetime value (LTV). LTV is typically conceived as a numerical value representing a customer's expected total contribution to company profits and is based on several measures:

1. the customer's expected lifetime, or tenure, with the company;
2. the per period (e.g., monthly) charges paid by the customer; and
3. the per period cost of the company's providing service to a customer.

These measures are mathematically combined, and often suitably discounted, for each individual customer to typically produce a single number, or LTV score, to be stored in the customer database. Scoring such as this has become very popular within services whose rates and service levels can be easily varied and is a central goal of the increasingly important company function of database marketing. This article offers a new way of evaluating customers, based on their dynamic relationship with the company. It describes a methodology for the estimation of the associated scores and shows how this can be used in company operations.

The basic building block of this methodology is a customer's hazard function, which is simply that customer's churn (exit) probability as a function of his or her age, that

The authors gratefully acknowledge the very substantial help of three anonymous referees and the editor.

Journal of Service Research, Volume 3, No. 3, February 2001 205-219

is, time with the service provider. Through this methodology—featuring a combination of statistical and data-mining techniques—estimation of an individual's hazard function, the main building block of this analysis, yields a variety of useful ways of guiding customer retention and targeting operations. In particular, we use the results of the lifetime analysis to suggest a method of differentiating company retention efforts based on our extension of LTV and of segmenting customers on the basis of their hazard functions.

Thus, this article proposes how the company manager can

1. develop a measure of the Gain in Lifetime Value (GLTV), an extension of traditional LTV that informs the company's attempts to target customers for retention;
2. more clearly understand customer relationships with the company and how they relate to commonly measured variables; and
3. segment the customer population in an operationally useful way.

THE VALUATION OF CUSTOMERS

In developing a relationship with a customer for a repeatedly purchased service, how should that customer be evaluated by the company? Important customers should be treated in special ways that enhance their profit production and increase the probability of their being retained. Customers with low value might be offered a service that is less costly to provide.

Simple revenue (or profit) has often been selected as a measure of customer value. "Big" customers are taken to be those who purchase a great deal of products or services, and these customers are often handled in special ways. This measure is indicative of the first level of understanding in valuation.

In an increasingly competitive world, however, it has been recognized that such "big" customers may be targeted by competitors or be fickle buyers in themselves so that their tenure, or period of custom, with a particular provider may be brief. A customer whose revenues per period are lower, but whose loyalty is greater, may be a better customer during a longer time horizon. This is the basic argument for incorporating the customer's estimated remaining lifetime in any valuation. Although we argue below that it too is only a partial measure, "lifetime value" (often abbreviated LTV or CLV, although we use the former here) has been considered to be the second level of understanding in valuation.

The notion of LTV has been developed to incorporate a customer's likely tenure with a particular service provider into the "simple revenue" method of valuation. Customer LTV has long been used in the mail-order industry to justify such marketing actions as the mailing of a specific sequence of catalogs (Schell 1990). The concept has also been applied in the industries of newspaper publishing (Keane and Wang 1995), retailing (Hughes 1996), insurance (Andon, Baxter, and Bradley 1998), and credit cards (McDonald 1998). More recently, LTV has been suggested as a basis for the design of marketing strategies; different segments of LTV scores would qualify customers for such special treatment as favorable pricing, customer service, equipment upgrades, or other concessions. Such a single-dimension measure may well be optimal for mail-order and other repeat buying behavior, but subscription services where customers have a well-defined birth and death time may require information of a different sort.

One major problem with traditional LTV is that it often presumes a fixed tenure distribution with a particular company; that is, the customer's tenure is implicitly assumed to be unsusceptible to increase by a company's actions. Consequently, a strict reliance on LTV might downgrade high-revenue customers whose estimated remaining tenure is low and upgrade loyal but low-revenue customers. Most relationship-oriented companies would not intuitively operate in this way.

Other authors have recognized that this classic notion of LTV thus gives little guidance to the company's retention managers. Strauss and Friege (1999) point out that "the LTV of the terminated relationship is not important. Only the value of the regained customer in the future is of interest" (p. 351). They go on to outline a method for evaluating the actions of customer retention or regaining given this anticipated LTV. Our work complements this analysis by offering an explicit methodology for calculating this fundamental building block. Bolton (1998) and Bolton and Lemon (1999) also consider how customers' lifetimes can change as a result of company interventions and thus affix a value for a general activity (e.g., increased training of the newly activated customer). We show below how to explicitly perform a similar lifetime increase calculation with individual customers for tactical retention activities.

We will show below that through the tenure concepts of LTV, an extension can be developed that incorporates company retention actions in a simple way and leads to a more reasonable targeting and relationship with intuitively important customers. We call this the customer's GLTV. Like classical LTV, this builds on the methodologies of survival analysis and, in particular, the estimation of the hazard functions of individual customers.

The following section is a brief overview of the basic ideas of survival analysis and the classical estimation of hazard functions. Further details are available in several standard texts (e.g., Cox and Oakes 1984).

SURVIVAL ANALYSIS BACKGROUND

A key concept that underlies our methodological approach is the classical statistical notion of a hazard function for an individual customer. This is a vector that displays the probability of that customer's churning—or "death" (of the relationship) in the language of survival analysis—at any given tenure (or "age"), conditioned on the customer's having already survived to that age. Mathematically, the hazard function for customer i is

$$\theta_i(t) = P(\text{death at age } t \mid \text{survival to age } t). \quad (1)$$

It is a slight simplification to assume that deaths only occur at integer times $t = 0, 1, 2, \ldots$, and the interval $[t, t+1)$ in which death actually occurred is indexed by t. The estimate of the hazard function for customer i is denoted by $\hat{\theta}_i(t)$.

An illustrative estimated hazard function is shown in Figure 1.

There are three classic statistical approaches for the analysis of survival data, largely distinguished by the assumptions they make about the parameters of the distribution(s) generating the observed survival times. All deal with censored observations by estimating hazard functions $\{\theta_i(t)\}$ where

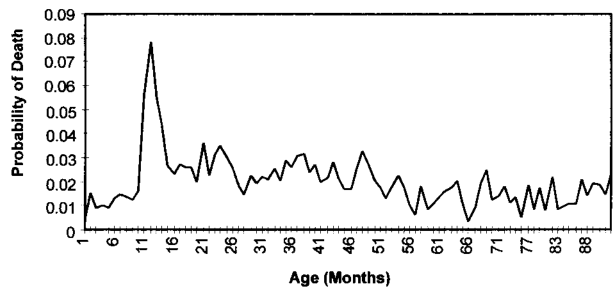
$$\theta_i(t) = \text{probability of subject } i\text{'s death at time } t \text{ given subject } i\text{'s lifetime is } t \text{ or greater} \quad (2)$$

or the survival function $\{S_i(t)\}$ where

$$S_i(t) = \text{probability that subject } i\text{'s lifetime is no less than } t. \quad (3)$$

Parametric survival models (e.g., Lawless 1982) estimate the effects of covariates (subject variables whose values influence lifetimes) by presuming a lifetime distribution of a known form, such as an exponential or Weibull. Although popular for some applications (especially accelerated failure models), the smoothness of these postulated distributions makes them inappropriate for our data which, as indicated in Figure 1, have built-in hazard "spikes" at times of contract expirations. Kaplan-Meier methods (Kaplan and Meier 1958) are nonparametric, providing hazard and survival functions with no assumption of a parametric lifetime distribution function. However, this popular estimator presumes a common hazard function for all members of a population (or prespecified population subsets). Finally, proportional hazards regression (Cox 1972) explicitly incorporates the effects of individual customer covariates, but at the expense of presuming a common hazard function of which each individual hazard function is a multiple. In our development below, the reader will see that each of these three models is overly restrictive for our data and that an artificial neural net (ANN) estimation technique yields more plausible hazard function estimates for each individual customer.

FIGURE 1
Typical Estimated Hazard Function



We reemphasize the centrality of correct estimation of an individual customer's hazard function. Once estimated, more immediately useful measures are readily available. An estimate of the median remaining lifetime of customer i at time t_0 is

$$t_i^* - t_0 \text{ where } \frac{\hat{S}_i(t_i^*)}{\hat{S}_i(t_0)} = 0.5, \quad (4)$$

in which the estimated survival function $\hat{S}_i(t')$ can be constructed from the estimated hazard function:

$$\hat{S}_i(t') = \prod_{t=0}^{t'} (1 - \hat{\theta}_i(t)). \quad (5)$$

Equally important, an individual customer's hazard function can be modified to reflect company retention actions, which, in turn, produce a notion of lifetime increase and LTV gain. The methodology for doing so is of great tactical interest to the manager in charge of customer retention.

FROM REVENUE TO LIFETIME VALUE TO LIFETIME GAIN

With the construction of hazard functions, and consequently estimates of an individual customer's survival functions and remaining lifetimes, we can now describe how to evaluate customers in terms of their potential reaction to company retention initiatives. It is well recognized in company operations that customers should be differentially valued. Limited resources must be targeted toward those customers who are central to the strategy of the business's development, and away from those of marginal (or even negative) worth. As we said earlier, some companies focus on "big" customers, where "big" is measured by the periodic (e.g., monthly) revenues they generate. In contrast, one could weight monthly revenues by the probability that the customer will survive (i.e., not churn) at future points in time, and this results in a measure of lifetime revenue. If company costs are subtracted at these future times, the resulting measure is often called LTV. This notion is attractive to company managers with a financial background, in that it answers the question, "What is a reasonable value for this customer as a current company asset?" It has been a popular guide to customer acquisition in the direct marketing activities of fund-raising and mail order. Note however, that LTV is intrinsic to an individual and does not account for any ongoing relationship a company has with its customers. In particular, a company's actions will often have an effect on his or her remaining tenure and future revenue generation. Strauss and Friege (1999) call this future profit Second Lifetime Value (SLTV). Although the activities of relationship marketing can be quite varied, we suggest a measure of GLTV below to quantify the potential financial effects of company retention efforts and thus to incorporate the company's actions on LTV. In sum, we distinguish three levels of customer valuation: (a) his or her single period contribution to a company's revenues, (b) his or her associated profit (revenue-cost) calculated during the entire lifetime of the relationship, and (c) the increase in that lifetime profit based on a company's efforts. Each level extends the previous level. Only the third, GLTV, incorporates company actions and thus has significance for retention operations and strategy.

Other authors have considered potential increases in customer lifetimes and their effect on managerial strategy. Using a rich set of customer covariates, including survey-based satisfaction measures, customer contacts, as well as usage and cost data, Bolton (1998) constructs a proportional hazards model for cellular telephone customer lifetimes. By examining the effects of changes in some of the covariates (e.g., customer satisfaction), she succeeds in estimating the effect on customer lifetimes of such strategies as increased regaining of new customers. Our aims here are arithmetically similar but managerially more tactical, focusing on the potential effect on lifetimes of contract renewal and rate plan changes for an individual. It becomes very important, then, to ensure that hazard functions are as disaggregated as possible, and the presumption of proportional hazards is relaxed of necessity through use of the ANN model. (Indeed, we demonstrate below that our customers can be segmented into several sets within which proportional hazards roughly holds, but whose baseline hazards differ markedly across segments.) Also, our need to construct a hazard function for each of the company's customers forces our use of a less rich covariate set from company databases than for Bolton's (1998) or Bolton and Lemon's (1999) situation, as survey data are only available for a limited set of customers.

Our situation and analysis proceed on the empirically based presumption that individual revenues and costs are essentially constant during the customer's lifetime. In many service industries, customer profitability will depend on his or her position in the service life cycle, and this can be built into the analysis below, although for simplicity we do not do so.

The neural network model produces individual-level hazard functions that are not bound by the proportional hazards assumption. From these estimated hazard functions $\{\hat{\theta}_i(t), t = 1, 2, 3, \ldots\}$ calculate the estimated survival function $\{\hat{S}_i(t), t = 1, 2, 3, \ldots\}$, as indicated by (5), with the estimated remaining lifetime at time t_0 given by (4).

Suppose for simplicity the monthly revenue generated by customer i is effectively constant for all t, with a value of R_i, then an estimate of the total expected revenue from customer i is

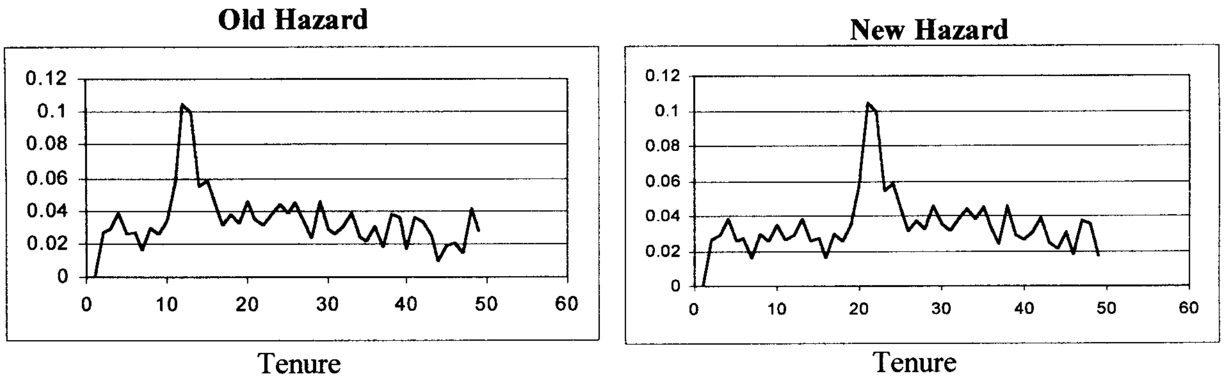
$$ER_i^0 = t_i^* R_i. \tag{6}$$

This quantity can be used to calculate an expected gain from a successful retention effort, and this, in turn, can be used to rank customers for targeting for special treatment.

Suppose customers are subject to a critical time period in their tenure when a decision is made concerning the company's efforts to continue the individual's patronage. This time, t_0, may be marked by the expiration of a contract or special promotion. Shortly before this critical time, say at time $t_0 - \Delta$, an attempt is made to restore the customer's potential future behavior to what was estimated at the beginning of his or her service period. That is, if this retention effort is successful, the customer's new hazard function is translated by the interval t_0. The illustrative graph in Figure 2 is based on a contract expiration date of $t_0 = 12$ months. When $\Delta = 2$ months (i.e., the retention effort is made 2 months before contract expiration), the customer's old and new hazard functions are shown in Figure 2.

Figure 2 illustrates only one way in which a customer's current hazard function might inform his or her transformed hazard function after a successful retention effort. Another plausible approach is to base a retention offer to a particular on some configuration of covariates (such as various usage

FIGURE 2
The Customer's Old and New Hazard Functions



or other service charges) and apply the estimated hazard function model—a function of covariates—to the new values. This will also result in a new hazard function.

For the ith customer, call this new hazard function $\{\theta_i^+(t), t = 1, 2, 3, \ldots\}$. This generates the corresponding survival function $\{S_i^+(t), t = 1, 2, 3, \ldots\}$. Using (4), one can calculate the median remaining lifetime at time t_0 for the new hazard function:

$$t_i^+ - t_0 \text{ such that } \frac{S_i^+(t_i^+)}{S_i^+(t_0)} = 0.5. \tag{7}$$

(This, incidentally, is not necessarily the same as the time period by which the hazard function is translated.)

Then estimate the expected revenue ER_i^+ via

$$ER_i^+ = t_i^+ R_i, \tag{8}$$

and GLTV for customer i from a successful retention effort is

$$GLTV_i = ER_i^+ - ER_i^0. \tag{9}$$

In this discussion we have made some assumptions to simplify the notation. Straightforward modifications may be made to this formulation in case revenues or costs per customer also change through the retention effort or in case monthly revenue is known not to be constant (e.g., to have seasonal variations).

This establishes the framework in which a customer can be evaluated and his or her potential contribution to future profit can be calculated. In the next section, we provide details on an ANN model to estimate the hazard functions that lie at the center of this valuation methodology.

AN ANN MODEL FOR INDIVIDUAL HAZARDS

As indicated earlier, the central technical issue in an LTV or GLTV estimation is the correct construction of hazard functions for individual customers. Classical proportional hazards regression is perhaps the standard technique in this regard but requires that all customer hazards be constant multiples of a baseline hazard for all, or at least for prespecified subsets of, customers. In many marketing applications, it may be implausible that such proportionality holds, and customer knowledge may be sufficiently limited so that any nonproportionalities cannot be discerned from information at hand. More generally, the analyst may doubt his or her ability to correctly specify the functions by which covariates combine to estimate the hazard multiples required for the regression aspect of the proportional hazards technique. Careful use of a special ANN network model addresses both potential problems. We will see below that this type of model dispenses with the proportionality assumption and does not require knowledge of the hazard multiple regression structure. Consequently, our ANN model is likely to produce better estimates of hazards and customer tenure than the classical approaches.

ANNs have become generally popular in the business world for estimating responses or scores based on a set of underlying covariates. Through intensive computation, they result in implicit functional relationships between customer covariates and response variables, and thus perform a task similar to a regression model. From a set of covariates $x_{i1}, x_{i2}, \ldots, x_{ij}$ from the ith customer, the ANN linearly applies a series of weights $\{w_{jk}^{(1)}\}$ for each of the k units in a "hidden layer" and set of a typically nonlinear activation functions

$$F_k\left(\sum_{j=1}^{J} w_{jk}^{(1)} x_{ij}\right),$$

where F_k is often chosen to be some sigmoid function such as an inverse tangent or logistic function. This results in a value at one (or more) "hidden layer(s)," and these values are transformed by other linearly applied weights and another activation function F to yield an output

$$F\left[\sum_k w_k^{(2)} F_k\right]$$

that is thus a highly nonlinear function of the covariates. One of several back-propagation algorithms adjusts the two sets of weights $\{w_{jk}^{(1)}\}$ and $\{w_k^{(2)}\}$ to satisfy some measure of agreement between the outputs and the observed responses from a training set.

These resulting functional relationships are usually too complicated to display and understand, so that model is made manifest through its predicted scores for the response variables, rather than through the regression coefficients and error measures familiar to a statistician. In particular, multilayer feed-forward neural networks (Haykin 1994) are nonlinear, universal-function approximators (Hornick, Stinchcombe, and White 1989) that have been used to estimate models too complex for ordinary linear or logistic regression. They are a natural technique to use as a flexible counterpart to a proportional hazards regression. See the above references for further basic information on neural net models.

What information is specifically needed to develop an ANN model of hazard functions? In the language of neural networks, we must identify a target attribute based on preclassified (or historical) examples. For instance, observed responses to a particular marketing campaign might be used to predict responses to similar future campaigns. Being a universal-function approximator, the neural network uses the preclassifed examples—termed *training data*—and learns to predict the target given a set of (independent) input covariates, without the need to prespecify a functional form for the predicting relationship. The process of constructing a neural net model with a single unbiased response variable is well-known. What makes the prediction of hazard functions different from now-standard ANN scoring applications is that (a) one simultaneously estimates a set of churn probabilities for many time periods in the customers lifetime, and (2) the target sets of churn probabilities must take into account the censoring and truncation often found in survival analysis problems. The way in which this is accomplished is described next.

We model a customer's target hazard using the following two attributes of customer data observed in a given time interval: (a) tenure, or age since service initiation (TENURE), and (b) a service termination flag (CHURN) indicating the formal ending of service during the time period. For simplicity, consider such data observed during one month, so that d_t is the number of terminating customers in that month; n_t is the number of customers at risk. Note that the definition of *customers at risk* is somewhat complicated when subjects are observed for multiple time periods. Then, the customer is in the risk sets for each of the ages he or she passes through during the observation period. To model customer hazard for the period $[1, T]$, for every record or observation i, we create a target vector $\{h_i(1), \ldots, h_i(t), \ldots h_i(T)\}$, with the following values (for $1 \le t \le T$):

$$h_i(t) = \begin{cases} 0 & 1 \le t \le TENURE \\ 1 & CHURN = 1 \;\&\; TENURE < t \le T \\ \frac{d_t}{n_t} & CHURN = 0 \;\&\; TENURE < t \le T \end{cases}. \quad (10)$$

The ratio d_t / n_t is the well-known Kaplan-Meier hazard estimator mentioned above. Intuitively, we set the target hazard to 0 when a customer is active, 1 when a customer has canceled, and to the Kaplan-Meier hazard if censored.

The vector $\{h_i(t)\}$ thus can be thought of as a raw hazard function for the ith individual that the neural network will relate to the underlying covariates. This approach is similar to Street (1998), except that we use the hazard function, instead of the survival function, as the desired output. This frees the neural network from having to constrain its outputs to be monotonic, as survival functions must necessarily be. Then, the parameters of the neural network are set up to learn probability distributions (Haykin 1994).

The ANN is completed by setting some fairly standard processes, among them being the choice of training, testing, and validation subsets of the customer data; the choice of the number and size of the ANN's hidden layers; the specific activation functions; and the model fit criterion. The ANN is then run and produces a hazard function for each individual in the data set.

Once the hazard function is estimated for each customer, a survival function can be estimated via (5) and the predicted median remaining tenure is given by (4). Thus, the output of the neural network model consists of a predicted tenure for each individual customer, as well as that customer's estimated hazard function.

Use of an ANN model thus has the important benefit of estimating an individual customer's hazard function, free of the restrictions of the classic statistical techniques.

However, the effective inscrutability of the resulting model indicates some caution in the use of its predicted hazard functions. Just as with neural nets that estimate single-response variables, there are some warnings to be issued:

1. The form of the function that relates covariates to hazard function components is not visible, as with a regression model.
2. The relative importance of each of the model's covariates also cannot be directly seen, as with regression coefficients.
3. Estimates of variances and other error measures are not generally available.
4. Problems with the data, such as actual or near collinearity of the covariates, or outliers, will not often be revealed, as it might by a regression or other classical tool.
5. Special features of the data, such as biased samples or correlated observations, are just as severe sorts of problems as they are for statistical analysis.
6. Results of the model fitting are not subject to intuitive expert judgment, as unexpected statistical parameters (e.g., incorrect coefficient signs) often are, although residual analysis is still available.

It follows that there are business activities that are not well supported by neural net predictions. Neural net models yield predictions but by themselves will not suggest possible business responses. In our situation, we may predict high postcontract churn for certain customers, but without further analysis, we cannot associate, say, an unfavorable rate plan with this pattern, which would suggest a retention tactic (e.g., offer a better rate plan). Thus, customer targeting is a viable ANN output, but suggesting a retention or other relationship tactic is not. ANNs are also not known for their stability. This can be an advantage in a changing population with unknown dynamics, where the ANN automatically generates a wholly new model. Modeling a slowly changing customer population at two points in time, however, can produce radically different ANN models, even though some factors are known to be static. Consider the introduction of a special offer—artificially low price for 3 months, with a large increase thereafter—in some small customer segment. Its effect would tend to dominate the refitting of an ANN, whereas a more controllable model—for example, a proportional hazards regression—could have the greater stability of the market built into its form. More generally, an ANN cannot easily incorporate subject matter knowledge, at least as compared with classical linear regression.

Consequently, our neural net modeling was preceded by a series of classical statistical models that indicated aggregate results on baseline hazards and lifetime residuals. The hazard segments produced by the ANN and the achieved residuals appeared to be intuitive generalizations of the preliminary statistical work and increased confidence in the full ANN results. On balance then, it is a great benefit to marketing strategy to have available the hazard functions of individual customers, unconstrained by the assumptions of classical proportional hazards modeling. Other researchers have noted the need for explicit disaggregated hazard functions, and some other techniques have previously been devised for their estimation. For example, Kooperberg, Stone, and Truong (1995) have developed a flexible hazard estimator using linear splines and their tensor products.

HAZARD MODELING APPLICATION AND RESULTS

The statistical and data-mining models discussed above can be applied to the customers whose information resides in internal company databases, producing valuable retention and other marketing information. The cellular telephone division of a major American telecommunications corporation has a customer data warehouse containing billing, usage, and demographic information. This configuration is common across many industries. The warehouse is updated at monthly intervals with summary information by adding a new record for every active customer and noting customers who have canceled service.

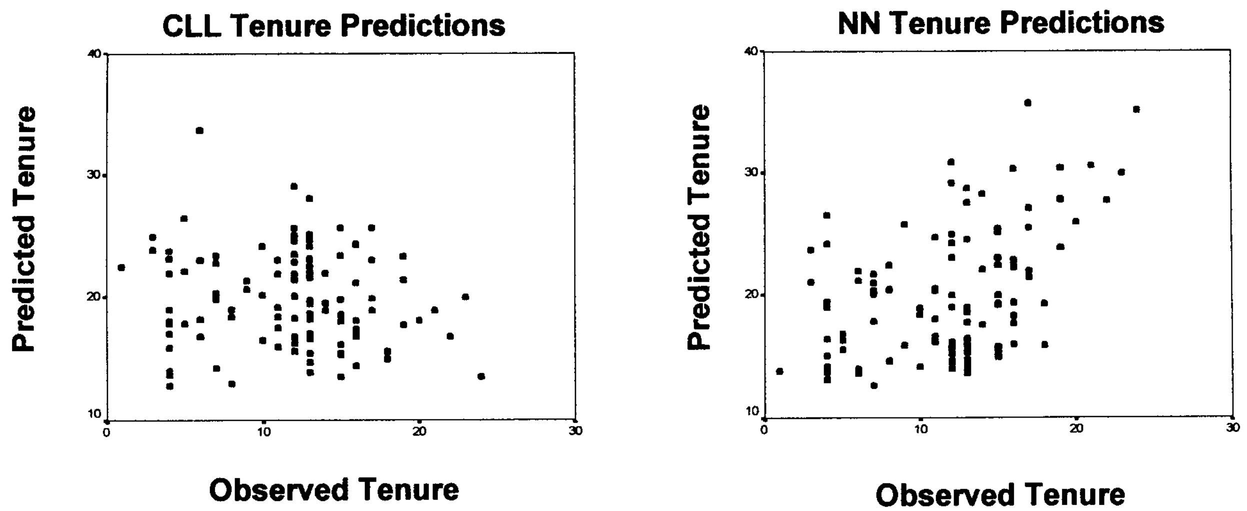
For the purposes of the tenure modeling reported in this article, we obtain a data extract from the warehouse where each customer record includes the following:

- *Billing:* previous balance, charges for access, minutes used, toll, roaming, and optional features
- *Usage:* total number of calls and minutes of use for local, toll, peak, and off-peak calls
- *Subscription:* number of months in service, rate plan, contract type, date, and duration
- *Churn:* a flag indicating if the customer has canceled service
- *Other:* age, current and historical profitability, optional features

These variables are related to those found in Bolton (1998) and Bolton and Lemon (1999), thereby allowing the relating of our modeled customer lifetimes to customer retention decisions through such constructs as customer satisfaction, payment equity, and the underlying utility of the service, as these authors have demonstrated.

We use a sample of data from a single relatively small market, comprising approximately 21,500 subscribers active during April 1998, and model their hazard functions from tenure $t = 1$ to $T = 36$. Thus, the churn variable in these data would indicate those customers who canceled service during the month of April 1998. (A customer be-

FIGURE 3
Complementary-Log-Log (CLL) and Neural Net (NN) Predictions



havior expert might legitimately argue that this definition of *churn* is strictly formal, based on a company decision rather than a customer decision. Passive customers, in particular, might signal their effective service termination by ceasing usage without formally notifying the provider, thus causing our observed lifetimes to be somewhat longer than behavioral lifetimes. Although this effect surely occurs, a separate analysis of our data showed no discernible drop-off in usage in the months preceding the formal churn date.) Only a small percentage of customers churned in that single period, so the vast majority of the observations are censored. Note that a straightforward examination of months in service (tenure) would yield means biased in two ways. First, the data are right censored, in that most customer's final lifetimes are not known as they have not yet churned. Second, the data are left truncated, in that the only customers observed are those whose lifetimes coincide with the observation period: short lifetimes in the population tend to be underrepresented in the sample. However, by basing our models on churn proportions among x-month-old customers (x = 1, 2, 3, . . .), both sources of these well-known biases are avoided. Both these remedies are now standard (see Allison 1995).

The classical proportional hazards model for the discrete death times we have here can be operationalized by its parameterization as a complementary-log-log (CLL) model—see Allison (1995) for more information including an explanation of what the CLL form is generally appropriate for the proportional hazards formulation. A comparison of one specific neural net and this proportional hazards model, both relying on the same set of potential covariates, is presented in Figure 3, which shows the graph of lifetime predictions versus observed lifetimes from the two models.

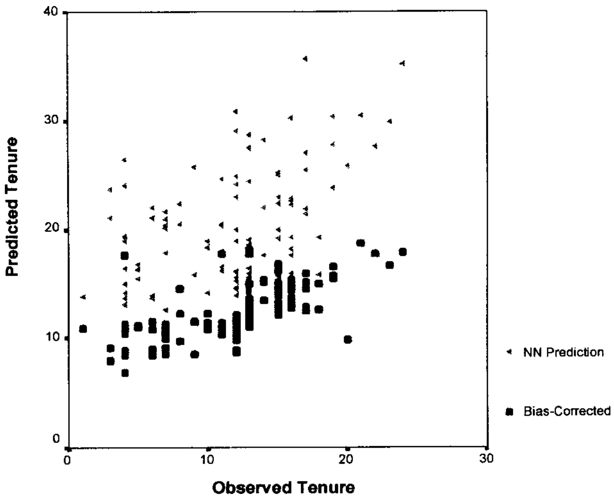
Figure 3 shows predicted median tenure at birth (as calculated by [4] for $t_0 = 0$) for the neural network and for the CLL model plotted against observed tenure for the 161 customers in the holdout data set who have already canceled. It is clear from the graphs that the neural network is more sensitive for predicting tenure than the CLL model. The CLL tenure predictions are clustered around 20 months and rarely exceed 30 months, resulting in a gross underestimate of tenure for long-lived customers. The neural network predictions, on the other hand, are more highly correlated with observed tenure, even for the higher tenure values.

The ANN's superiority for these data is also shown by the residual plot in the appendix. Although we do not display the evidence here, similar results have been seen for several different cellular markets at several different points in time. It turns out that a major source of the ANN's superiority for these data is its ability to discern distinct clusters of hazard functions, each with a unique baseline hazard.

Both the CLL and neural network estimators, however, give biased tenure estimates. An analysis of its cause and an approximate correction are given in the appendix. Note that the best-known causes of bias, namely, right censoring and left truncation, have already been accounted for and are not the source of this problem. The bias-corrected version of the neural network estimator and the original neural net-based predictions yield the plot presented in Figure 4, showing predicted tenures versus observed tenures.

It is evident that the bias-corrected estimate gives a more accurate prediction for the ranges of tenures most

FIGURE 4
Bias Corrected Versus Observed Tenure



NOTE: NN = neural net.

frequently observed in these data (7-20 months) than the original ANN prediction, which, in turn, is significantly better than the proportional hazard predictions. The bias in these estimates still shows some dependence on actual tenure, but it is less pronounced than for the ANN predictions.

In sum, these data suggest that the data-mining technique of ANNs are superior to classical statistical models for predicting customer tenure. They also have the advantage of producing a separate hazard function for each customer that is not necessarily proportionally related to other customers' hazard functions. However, successful tenure estimation is not fully automatic: A Kaplan-Meier–like hazard estimator must be borrowed from classical statistics to form the target hazards, and then a special regression is necessary to remedy the bias.

CUSTOMER SEGMENTATION BASED ON HAZARD FUNCTIONS

A very important use of individual hazard functions as produced by an ANN is the calculation of GLTV. However, the hazard functions are themselves interesting and yield important marketing insights when analyzed by more traditional statistical methods. The distinguishing feature of the ANN modeling described above is its production of hazard functions for each customer. Just as marketing insight flows from the segmentation of customers from their vectors of attributes, it is worthwhile to cluster the T components of these hazard functions. As shown below, segmenting these hazard functions gives important operational guidance for customer targeting and retention efforts.

The modeling was based on 21,500 subscribers, of whom 15,000 were used to construct a hazard function model, and the rest was used as a holdout sample whose results we focus on here. These approximately 6,500 hazard functions produced by the neural network model are clustered as follows. Each hazard function then consists of T = 36 components representing churn probabilities at customer ages 1, 2, . . ., 36. Construct a small number of statistics to indicate the shape of each hazard function for each customer. These include the overall slope of the hazard function from 1 to 36 months, the relative size of any "spike" at the contract expiration time of 12 months and of 24 months, and so on. This results in a small number of statistics for the hazard function of each customer, based simply on the hazard function's shape. Then k-means clustering (10 clusters maximum, assimilation of small-size clusters into larger ones) is used to assign a cluster membership to each customer.

Because each cluster contains up to several thousand customer hazard functions, a special technique was needed to display the hazard functions within each cluster. First, the 36 components of each hazard function were analyzed by principal components. This technique (see, e.g., Morrison 1967) essentially reorients the 36 dimensions of the hazard components with new axes in the hope that some small number of these new dimensions will capture most of the variation of the original data. By convention, the first principal component is labeled so it captures the most possible variation. In our data, we arrange the hazard functions along this first principal component and then display the hazard functions that fall along regular intervals in this new dimension. In Figure 5, the intervals are chosen to be the 10th, 25th, 50th, 75th, and 90th percentiles. This range of hazard functions shows the variety of shapes and spread within a cluster. The cluster proportions are shown alongside their displays.

Figure 5 suggests a fact confirmed by the principal components analysis: Nearly all variation within a cluster is captured by the first principal component, and therefore the hazard functions within a cluster are all nearly multiples of each other. Thus, a proportional hazards model roughly holds within each segment. These multiples are a way of arraying the hazards within each segment, and Figure 5 shows the segment hazards with regularly spaced functions displayed.

These four hazard clusters constitute a useful customer segmentation for the service operations manager. This segmentation has important interpretations of a customer's state of mind in using cellular service and important intuitive implications for the company's retention efforts for these different segments. These are summarized in Table 1.

FIGURE 5
Hazard Clusters

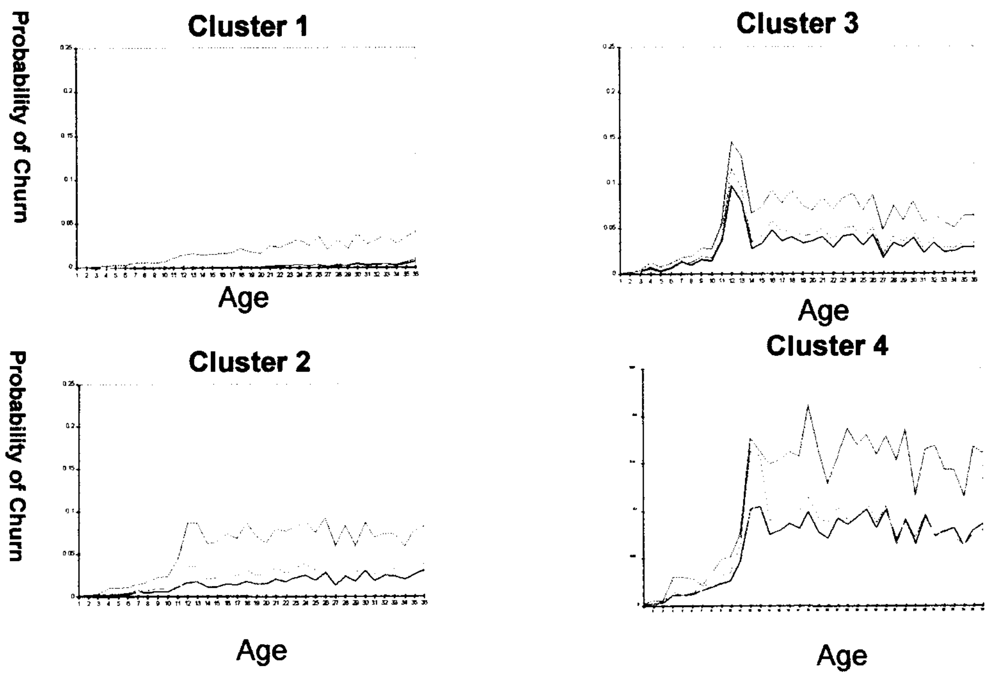


TABLE 1
Marketing Implications for Hazard Clusters

Cluster Segment	*Contract-Related Exit Timing*	*Implications for Retention Effort*
1	No effect of contract expiration	No preexpiration contact required; contact may trigger churn
2	Small increase in churn propensity at expiration; postexpiration churn propensity remains elevated	Moderate preexpiration effort needed; new contract or continued contacts needed
3	Large spike in churn at expiration; low churn thereafter	Concentrate effort on preexpiration; contract renewal may not be required
4	Large increase in churn at expiration; postexpiration churn high and increasing	High-intensity preexpiration effort; continued competitive offers to designated customers

Thus, these patterns have an important meaning for company marketing and retention efforts. Insofar as these patterns are discerned by a neural network model, whose mechanisms are effectively unknowable, and because the clusters are based on mere geometric shapes of the resultant hazard curves, company needs require their relating to internal data and customer histories. Although a classical tool such as discriminant analysis could be used for this task, the likely clusterwise differences in explainers and their potential nonlinear effects indicated use of decision tree (recursive partitioning) methods (Safavian and Landgrebe 1991). In this method, the entire data set, with its initial distribution of the four cluster types, is repeatedly split based on values of explanatory variables. In an ideal analysis, the splitting results in subsets of data that consist solely of one cluster or another. In our data, the explanatory variables included the following:

Detailed billing	A special, extracost feature; often associated with business customers
Total charges	Total charges on a customer's monthly bill; includes access charge, air time, roaming charges, and so on
Peak MOU	Number of monthly minutes of use (MOU) billed at defined peak hours
Channel	Sales channel (e.g., GTE distributor, auto dealer) through which the service was initially purchased
Total calls	Total number of calls in a month

Of course, many other variables are actually available for use in this analysis; the table above only highlights those that have an important effect on identification of the hazard function clusters. They are comparable to the covariates found in a study of a similar service in Bolton and Lemon (1999). Note that our implicit goal is the relation of a customer's retention decision, here characterized by hazard cluster membership, with the role these specific covariates may play in that decision. This is a necessary step in using the ANN-produced hazards to suggest possible retention strategies and tactics a company might consider.

The rules that define paths to the most discriminatory subsets are summarized in Table 2. Intuitively, it seems that Cluster 1 is composed of customers who are insensitive to contract expiration. This might include the "safety and security" set, who possess their cellular telephone as an emergency and convenience device, although it is of managerial interest to note that there are many high-revenue customers in this cluster. Cluster 3 comprises users who have a moderate flat-rate access charge that accommodates all their calling needs. Cluster 4, in contrast, comprises customers with rate plans whose flat rates do not fit their high calling volumes. These may well be customers who would be better served by a different rate plan; their high postcontract churn probabilities indicate that such improved plans often may be obtained through alternative suppliers. It may be that Cluster 2, which is a scaled-down version of Cluster 4, may also include customers with inappropriate contracts.

In these data, it is likely that the expiration of 12-month contracts plays a major role in determining the shape of the hazard functions and the clusters that they fall in. (It is possible that the spikes of some customers could theoretically be due to some other influence, such as a seasonal effect like Christmas or an informal 12-month "tickler" reminder set by the customer himself, but the choice of month—April—for this analysis gives no support to this notion. Contract expiration is known from exit interviews to be a major trigger of churn.) Note, however, that meaningful hazard function construction and segment interpretation do not require the existence of contracts. Generally, this sort of analysis helps to understand churn at different stages in a customer's life cycle. Cluster 1 above is interesting for the relative constancy of low churn probabilities across the constituent customer's lifetimes. Cluster 4 is characterized by consistently high churn probabilities after an initial quiet period. Both findings would be interesting and operationally useful even if there were no 12-month contract to expire.

Indeed, this last observation suggests that the production, clustering, and analysis of hazard functions are useful, even when there is no single event such as contract expiration to visually dominate the hazard form. The discovery of segments with high absolute churn rates (as Cluster 4 here) or declining hazards (as postexpiration Cluster 3) have obvious retention and related marketing consequences. We have suggested here that relative goodness of the customer's rate plan is associated with the declining hazard rate in Cluster 3. In general, then, our methodology can yield insights even when contracts and their expirations are not an issue.

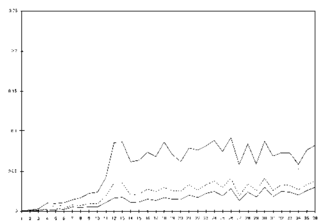
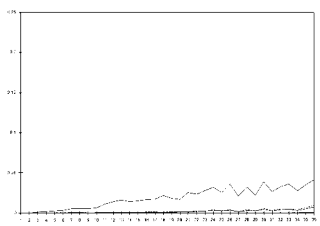
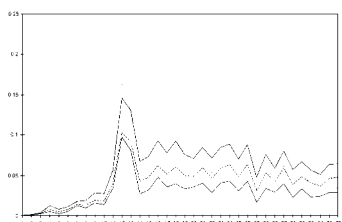
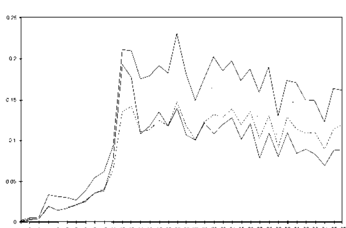
AN EXAMPLE FOR THE USE OF THE GLTV ESTIMATE

We have shown that the GLTV concept developed in equations 6 through 9 has various operational uses. It serves as a guide for the company's interactions with individual customers when their retention may depend on a modified pricing plan or other concessions. It also becomes the basis for forming customer valuation groups toward which different retention efforts and concessions might be offered. In many large companies, it is simpler to administer a retention program in which company actions are directed toward segments of customers, rather than individuals. In this section, we describe how GLTV has been used to construct relationship segments and how that segmentation compares with traditional segments based on revenue or on classic LTV.

Suppose, then, that retention efforts are varied by some measure of customer valuation, such as revenue, LTV, or GLTV. Segments are constructed on the basis of one of these measures of customer value, so that the highest $100f_1\%$ are subject to one (high) level of retention effort, the next $100f_2\%$ are subject to a different (and slightly lower) effort, and so on until the lowest $100f_s\%$ receive yet another (very low or nonexistent) retention effort. A simple and common case is when $S = 3$, so that the highest $100f_1\%$ are given so-called premium treatment, the bottom $100f_3\%$ receive a minimal treatment, whereas the remainder receives the company's standard retention effort. Possible generic retention strategies have been cited elsewhere in the customer retention literature. See, for example, Zeithaml, Berry, and Parasuraman (1996) or Bolton and Lemon (1999).

The efficacy of this segmentation is calculated as follows. Suppose customers are segmented by *GLTV*. For a retention effort for the *i*th customer in the *s*th segment, there will be a certain probability that the customer will choose to renew his contract. Call that probability p_s. Note that it is intended that $p_1 = 0$, or some small probability. Let $\overline{GLTV}_s$ be the mean *GLTV* of those customers placed in the *s*th segment. Then, the total expected gain from this segmentation is

TABLE 2
Characterizations of Hazard Clusters

Cluster Number	*Shape*	*Distinguishing Features*
2		A reference shape
1		Detailed bill; few calls/month
3		Two distinct types: • Zero charge for minutes of use (MOU), many calls/month • No detailed bill, low total charge
4		High total charge

$$N = \sum_{s=1}^{S} f_s \, p_s \, \overline{GLTV}_s, \tag{12}$$

where N is the total number of customers in the population of interest. This calculation allows the comparison of this general retention strategy, and its basic segmentation, with other candidate segmentation methods (e.g., revenue or LTV). Assume all candidate segmentations construct the same size segments $\{f_s, s = 1, 2, \ldots, S\}$ so that the segmentations differ only in the ordering of customers and therefore in the configuration of customers who are assigned to a given ~~segmen~~t. For a given candidate segmentation, calculate $\{\mathrm{GLTV}'_s, s = 1, 2, \ldots, S\}$, the set of mean *GLTV*s for this segmentation, and further calculate the total mean expected gain from this segmentation.

$$N = \sum_{s=1}^{S} f_s \, p_s \, \overline{\mathrm{GLTV}'}_s. \tag{13}$$

The difference between (12) and (13) is the dollar amount by which a candidate segmentation basis is different from the one based on GLTV.

To illustrate the operational effect of this proposed segmentation, we consider two relevant alternatives, much like what has been done for real cellular telecommunications data. One might form segments based on average monthly revenue, or on a measure of lifetime revenue. (Of course, in many situations, one might also consider the cost of service provision and extend these two bases to form a profitability index or lifetime profit.) We choose the simpler formulation as it better illustrates the effect of the three notions of customer value discussed at the beginning of this article.

We illustrate these segmentation effects with the previously mentioned sample of approximately 6,500 customers from an actual telecommunications market, comprising roughly a quarter of a million customers (i.e., $N \approx 250{,}000$). Each segmentation basis produces a particular ordering of this company's customers. The segmentation scheme relegates the bottom 10% to an intentionally reverse-incentive segment, 80% to a "business-as-usual" or default segment, and selects the top 10% for membership in a premium service segment. As above, let p_1, p_2, and p_3 be the probabilities of successful retention in each respective segment. Calculate the average GLTV for each segment for the three segmentation method discussed above.

Table 3 shows these values (in units linearly transformed from dollars to preserve company propriety) for the three segments (top 10%, middle 80%, bottom 10%) for each of the three segmentation methods (monthly revenue, lifetime revenue, lifetime gain).

By construction, the GLTV method must have the lowest possible values for the bottom 10% and the highest possible values for the top 10%. The key point to note is that the LTV means are very different from those for GLTV, largely because many in the top 10% for GLTV have short estimated tenures, which decreases their LTV score. As we noted near the beginning of this article, many of these high-revenue/short-tenure customers could be well worth retaining, but their relatively small LTV scores would condemn them to a lower segment through that short tenure.

From previous studies in this market, it is estimated that $p_2 = .08$ and $p_3 = .35$, and it is presumed that p_1 is approximately 0. The difference between (12) and (13) can then be calculated to show the monetary effect of the GLTV segmentation compared with segmentation based on revenue. With these numbers, we calculate that the GLTV segmentation results in a gain over LTV-based segmentation of $3,602,250. This market is a relatively small one. For larger markets, the gains are proportionately larger.

TABLE 3
***GLTV* for Three Segmentation Bases**

	Segmentation Method		
Segment	*Revenue*	*Lifetime Revenue*	*Gain in Lifetime Value (GLTV)*
Bottom 10%	145.542	157.174	132.011
Middle 80%	375.579	435.927	373.520
Top 10%	1,739.133	1,218.467	1,744.266

SUMMARY

We have described and contrasted common notions of customer valuation and considered their description by scores of several kinds. Some important marketing information is obscured by reliance on any single score, but an extension of LTV, called GLTV here, can incorporate the effect of company actions on valuation and therefore guide company concession tactics and segmentation strategies. This concept is most effective when individual customers' hazard functions can be estimated, and this can easily and meaningfully be done with a special neural net model. In this article we have argued that

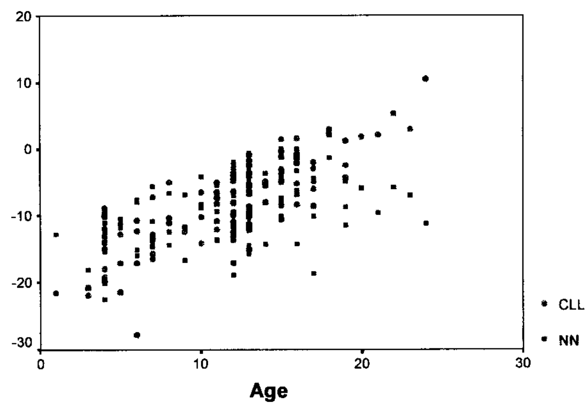
1. customer valuations by revenue and LTV ignore the potential effects of company actions, particularly retention and service actions;
2. valuations based on an individual customer's hazard function provide meaningful customer segments and a method of modeling the profit effects of company retention activities;
3. hazard functions can be freed from the restriction assumptions of the classical statistical models (e.g., proportionality) by constructing an artificial neural net;
4. hazard functions display information about a customer's service subscription lifestyle and suggest a customer's potential for a gain value under successful subscription retention;
5. the effect of a successful customer retention effort can be quantified with these hazard functions, thus quantifying lifetime gain (GLTV) in lifetime revenue;
6. GLTV can be used to segment customers and, in the data analyzed here, substantially improves the profitability of company retention operations; and
7. the different retention behaviors indicated by the ANN-based segments indicate very different marketing strategies among those segments.

Thus, it appears that GLTV and the individual, nonproportional hazard functions supporting it can significantly improve company operations in the areas of customer valuation, marketing strategy, and retention targeting.

APPENDIX
Correcting for Bias

Even though the estimates of customer lifetimes shown in Figure 3 seem reasonable, a close examination shows that the tenure estimates are systematically larger than observed lifetimes, both for the complementary log-log (CLL) and the neural network model. Figure A1 shows residuals (i.e., observed minus predicted tenures) for the CLL and artificial neural net (ANN) models, plotted against observed tenures.

FIGURE A1
Residuals: Neural Net (NN) and Complementary-Log-Log (CLL)



Not only are both sets of lifetime estimates biased, but the bias appears to be dependent on observed lifetimes. Why do both models overestimate tenure, even though it is known that the CLL model asymptotically estimates the hazard rates without bias? Let h_t be the estimated hazard at time t for a given customer, and let $\theta_t = E(h_t)$. Recall from (10) that customers' expected lifetime is given by

$$\tau \text{ such that } 0.5 = \prod_{t=0}^{\tau}(1-\theta_t)$$

or, equivalently,

$$\tau \text{ such that } \ln(0.5) = \sum_{t-0}^{\tau}\ln(1-\theta_t)\,. \qquad \text{(A1)}$$

It follows from Jensen's Inequality that $E(\ln(1-\hat{\theta}_t) \leq \ln(1-\theta_t)$, $t = 0, 1, 2, 3, \ldots$ Furthermore, the proportional hazards structure, which holds for all customers in the CLL model, and within clusters for the ANN model, suggests that, approximately,

$$\frac{1-\hat{\theta}_t}{1-\theta_t} = \delta, t = 0,1,2,3,\ldots$$

Then it is possible to approximate a form for the bias in the ANN and CLL estimates, and hence to correct for it. Suppose t* is the estimated tenure for a given customer, that is,

$$\ln(0.5) = \sum_{t=0}^{t*}\ln(1-\hat{\theta}_t). \qquad \text{(A2)}$$

Subtracting (A1) from (A2),

$$0 = \sum_{\tau}^{t*}\ln(1-\theta_t) + \sum_{t=0}^{t*}\ln\left(\frac{1-\hat{\theta}_t}{1-\theta_t}\right). \qquad \text{(A3)}$$

Expanding in a Taylor Series, and observing in our data that the hazard rates are quite constant somewhat past the median lifetime, the first term of (A3) is roughly $(t^*-\tau)\bar{\theta}$, where $\bar{\theta}$ denotes a mean hazard rate, whereas the second term is approximately $t^*\ln(\delta)$. Thus, (A3) becomes

$$t^* - \tau = \frac{\ln(\delta)}{\bar{\theta}}, \text{or } \tau - t^* = -\frac{\ln(\delta)}{\bar{\theta}}.$$

To use this form, in the data at hand, use the training and test data sets from the neural network to fit

$$\frac{pred - obs}{obs} = d\left(\frac{1}{\bar{\bar{h}}}\right)$$

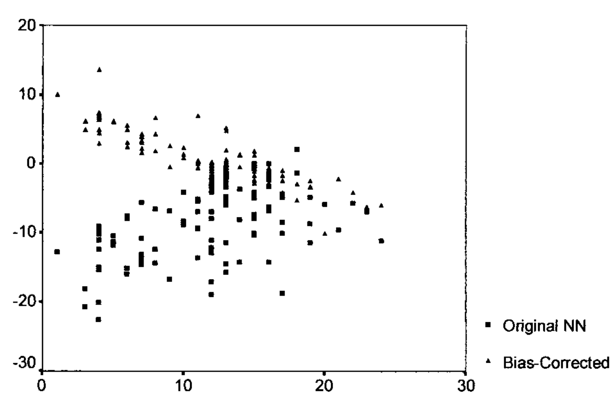
across all customers who have churned, with *obs* being their actual observed tenure, *pred* being their tenure predicted by the neural network model, and $\bar{h}$ being the mean estimated hazard for the last few (10 months was chosen here) months of the customer's lifetime. Then the bias-corrected estimate of customer tenure is given by

$$pred' = pred + \hat{d}\left(\frac{1}{\bar{\bar{h}}}\right),$$

where $\hat{d}$ is the estimate of d from the regression above.

Form residuals from both *pred″* and *pred* by subtracting the observed tenure from those customers in the validation set who have churned. The graph of these two sets of residuals, plotted versus actual tenure, is shown in Figure A2.

FIGURE A2
Residuals: Original Neural Net (NN) and Bias Corrected



The bias-corrected predictor is closer to 0 on average than is the uncorrected neural network predictor, and its relation to observed tenure is less pronounced.

REFERENCES

Aaker, David, Vijay Kumar, and George S. Day (1998), *Marketing Research*. New York: John Wiley.

Allison, Paul D. (1995), *Survival Analysis Using the SAS® System*. Cary, NC: SAS Institute.

Andon, Paul, Jane Baxter, and Graham. Bradley (1998), "The Calculation of Customer Lifetime Value (CLV): Theory and Practice," in *Proceedings of the EIASM Workshop on Quality Management in Services VIII*, B. Strauss, ed. Ingolstadt, Germany: Catholic University of Eichstaett, 699-743.

Bolton, Ruth (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.

——— and Katherine Lemon (1999), "A Dynamic Model of Customer' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction," *Journal of Marketing Research*, 36 (May), 171-86.

Cox, David R. (1972), "Regression Models and Life Tables,"*Journal of the Royal Statistical Society*, B34, 187-220.

——— and D. Oakes (1984), *Analysis of Survival Data*. London: Chapman and Hall.

Haykin, Simon (1994), *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.

Hornick, Kurt, Maxwell Stinchcombe, and Halbert White (1989), "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2, 359-66.

Hughes, Arthur M. (1996), *The Complete Database Marketer*, rev. ed. New York: McGraw-Hill.

Kaplan, E. L. and Paul Meier (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457-81.

Keane, Timothy J. and Paul Wang (1995), "Application for the Lifetime Value Model in Modern Newspaper Publishing," *Journal of Direct Marketing*, 9 (Spring), 59-66.

Kooperberg, Charles, Charles J. Stone, and Young K. Truong (1995), "Hazard Regression," *Journal of the American Statistical Association*, 90 (429), 78-94.

Lawless, Jerald F. (1982), *Statistical Models and Methods for Lifetime Data*. New York: John Wiley.

McDonald, William J. (1998), *Direct Marketing—An Integrated Approach*. Boston: McGraw-Hill.

Morrison, Donald F. (1967), *Multivariate Statistical Methods*. San Francisco: McGraw-Hill.

Safavian, S. Rasoul and David Landgrebe (1991), "A Survey of Decision Tree Classifier Methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, 21, 660-74.

Schell, Ernest (1990), "Getting to the Bottom of Lifetime Value," *Catalog Age*, 7 (8), 67-70.

Strauss, Bernd and Christian Friege (1999), "Regaining Service Customers," *Journal of Service Research*, 1 (4), 347-61.

Street, W. Nicholas (1998), "A Neural Network Model for Prognostic Prediction," in *Proceedings of the Fifteenth International Conference on Machine Learning*. San Francisco: Morgan Kaufmann, 540-46.

Zeithaml, Valerie, Leonard L. Berry, and A. Parasuraman (1996), "The Behavioral Consequences of Service Quality," *Journal of Marketing*, 60 (April), 31-46.

James H. Drew is a principal member of technical staff in the Analytical Decision Support Department of Verizon Laboratories. He has a B.A. in mathematics and philosophy from Williams College, a diploma in mathematical statistics from Cambridge University, and a Ph.D. in statistics from Iowa State University. He is a member of the American Statistical Association, a fellow of the Royal Statistical Society, and a chartered statistician.

D. R. Mani is a principal member of technical staff in the Analytical Decision Support Department at Verizon Laboratories and a member of the Knowledge Discovery in Databases (KDD) project. Prior to joining Verizon Labs, he was a principal research engineer at Thinking Machines Corporation involved in the design and implementation of parallel data-mining algorithms. He has a bachelor's degree in electronics and telecommunication engineering from Bangalore University, India; a master's degree in computer science from the Indian Institute of Technology, Kanpur; and a Ph.D. in computer science from the University of Pennsylvania, Philadelphia. He was recently elected senior member of the Institute of Electrical and Electronics Engineers (IEEE) and is also a member of the Association for Computing Machinery (ACM) and the American Association of Artificial Intelligence.

Andrew L. Betz is a principal member of technical staff in the Analytical Decision Support Department at Verizon Laboratories. He has a B.A. in psychology from Bowling Green State University, an M.A. in social psychology from Ohio State University, and a Ph.D. in social psychology from Ohio State University. He is a member of the American Marketing Association and the Psychonomic Society.

Piew Datta is a principal member of technical staff in the Analytical Decision Support Department at Verizon Laboratories and a member of the Knowledge Discovery in Databases (KDD) project. She received her B.S. in information and computer science from the University of California, Irvine, in 1989; her M.S. in 1992; and her Ph.D. in 1997. She is currently a member of the American Association of Artificial Intelligence, ACM, IEEE, and the Society of Women Engineers.