



SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN



WORKING PAPER SERIES

Forecasting Customer Lifetime Value

by

Joseph Ryan G. Lansangan¹
and
Erniel B. Barrios²

UPSS Working Paper No. 2010-12
September 2010

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

¹ Assistant Professor, School of Statistics, University of the Philippines Diliman

² Professor, School of Statistics, University of the Philippines Diliman

Abstract

In a highly competitive sector offering fast-moving consumer goods and services, recruitment cost is generally higher than retention costs. It is important then for marketing planners to develop strategies of attracting customers and engage them to patronize the products. One common strategy used is to lock them up for a specific period to patronize the product with a minimum usage. The customer can only be motivated to enter into such a contract if attractive incentives are available. The value of the offer along with the duration of the lock up period is determined based on lifetime profitability of the customer. Using the customer usage database, sampling, segmentation, modeling, and validation techniques in data mining are used to develop models that can forecast the customer lifetime value. The highly varied customer database being mined allowed the inclusion of uncertainty components in the estimation of customer lifetime value. A hazard function model based on a truncated lifetime data of customers can provide enough information to compute the value of the incentive to be offered and the length of the lock up period for the customer.

Keywords: data mining, customer lifetime value, customer relationship management, truncated data, hazard function model

1. Introduction

The dynamics of customer valuation expand with the growth of non-static data together with the profile (possibly, static) data. Most companies consider customer valuation by looking at customer behavior based on (internal or external) transaction and/or experience data, or from perception and/or opinion studies. Ideas such as how likely, how much, how many, or how soon can be inferred from understanding customer behavior – how likely he/she will discontinue engagement, how much profit is expected, how many he/she will possibly influence, or what or how soon the next purchase is expected. This information is valuable input in customer relationship management. In a credit card company, annual dues can be waived provided that the customer can be profitable within the period. A subscriber in a telecommunication provider can be offered incentives like equipments of discounts provided that patronage for certain periods can be assured. Similar is true for privilege card for restaurants and shops, mileage card for airlines, etc.

With the volume of data at hand, data mining techniques are commonly employed to extract the patterns useful to achieve the company objectives. Classification and/or clustering methods are used to identify “natural” groups based on data. Most methods draw on association rules to discover relations among different purchases and come up with likely “bundles” or “basket mix.” Regression models are developed to identify possible factors of an event of interest (e.g. defaulting, acquisition/purchase, etc.). Indeed, companies take advantage of many statistical techniques and procedures paralleled to their business objectives. As such, customer life cycle management remains as among the priorities of most if not all companies, having and/or instituting units such as in business intelligence (BI), customer relations/engagement (CR/CE), or market research (MR) groups.

Customer lifetime value (CLV) modeling is at the core or parallels to customer life cycle management. Different approaches are carried out to understand if not assess customer “importance” in relation to the company (or profit/earnings). A “score” is calculated for each customer, reflecting “how much” a customer is worth or equivalently how much the company should spend to retain the customer (or equivalently, acquire a new one). Among the more

common approaches, or as input to CLV itself, are customer segmentation, modeling customer average revenue or return on investment (ROI), customer longevity (i.e. occupancy period, engagement measured through time, or tenure), termination/churn rate, and acquisition/retention cost.

We develop customer lifetime value models using a database on customer usage database of a service industry company. Transaction details stored in its data warehouse (recording at a customer-level) are used with variables like amount/cost of service, count/frequency of purchase of service, and duration of service access. Customer account details are also kept, with information for the type of customer, customer's sex, age, date when first purchase started (equivalently, engagement duration or tenure), and as to whether the customer is continuing (or active) or non-continuing (historic).

As a metric of customer's worth, the company's CLV scoring were based on the customers' usage and/or patronage of its products/services. It is imperative for the company to assess its ROI on a particular customer, and along side foresee the customer's "engagement" measured in terms of tenure with the company. Thus, modeling efforts focused on looking at both average revenue and tenure.

2. Customer Lifetime Value: Concepts and Uses

From a supplier-oriented perspective, (Bauer, et al, 2003) defined customer value as the customer's economic value to the company and proposed to use customer lifetime value in corporate valuation. Customer activities contributing to company income like up selling, cross selling, referrals, etc., can benefit from using customer value as benchmark of profit computation and resource allocation.

Many frameworks were proposed in the measurement of customer lifetime value (CLV) as they are linked to other activities that stimulate corporate dynamics. (Berger, et al, 2006) proposed the chain of effects framework in the understanding of the influence of CLV on shareholder value. Just like in other frameworks, measurement of CLV is a major issue. They suggest that since CLV measurements are usually based on past data, elements of uncertainty should be incorporated in the models used to measure it and that expectations rather than deterministic accounting measures should be used.

Customer relationship management is an important component of business intelligence. It provides an integrative framework in which marketing activation and customer activities collaborate to increase patronage. (Venkatesan and Kumar, 2004) noted that CLV can also be maximized within the framework that allows managers to maintain/improve customer relationships through marketing contacts across channels. They further suggest that CLV is a viable metric for customer selection and marketing resource allocation. (Malthouse and Blattberg, 2005) emphasized that firms can be more profitable if they identify the most profitable customers (with high CLV) and invest favorably large among of marketing resources on them rather than those in lower expected CLV. (Dwyer, 1989) noted the value of CLV in support marketing decision-making. CLV and consumer behavior can be jointly featured in planning for optimal marketing mix that will lead to the improvement of long-term profitability.

Among the early proposal to measure CLV, (Berger and Nasr, 1998) used a series of mathematical models based on theory and on assumption grounded in customer behavior. Some of the possible applications of CLV from this process includes: relationship marketing,

marketing and promotional resources allocation, assessment of the effect of marketing strategy in acquisition and retention rates, costs and tradeoff, and customer equity.

To maximize outcomes of certain marketing activities on CLV in a telecommunications company, (Rosset, et al, 2003) suggests the development of CLV model based on segments of the population of consumers and model customer value and tenure separately before the integration later to come up with CLV. As noted by (Lee, et al, 2006), valuation of customer based on financial indicators alone could be misleading since there are other latent factors that can potentially influence CLV. Further highlighted that managing customer relationship by inputs (acquisition and retention) and outputs (revenues) for each customer can facilitate better prioritization of marketing investments by differentiating customers. (Ryals and Knox, 2005) suggest to measure risk-adjusted CLV and its impact on relationship marketing strategies and shareholder value.

There are several works on the uses of CLV in marketing and strategic business decisions. (Gupta, et al, 2006) postulate that companies increasingly generate income by initiating and enhancing a sustainable relationship with the customers. Incentive planning using CLV was proposed by (Reinartz and Kumar, 2003) as a strategy to increase profitability of customers

3. Methodology

Stratified sampling from the customer base was employed, based on status (active or historic), account type (corporate, organization/retailer, personal), and tenure cuts (*recent* or less than 1 year, *short term* or 1-2 years, *mid-term* or 2-3 years, and *enduring* or more than 3 years).

For the tenure analysis, two different models were developed, one which used only tenure thru a survival model (i.e. hazard functions) for expected lifetime (EL), and the second which used a generalized regression model with (left and right) censoring to estimate end-of-lifetime (EOL) after current period. Under the first approach, hazard functions are fit on the data set and on each of the pre-determined groups based on customer's age group, gender, account type and customer tag (regular or premium). Under the second approach, transactional data were accumulated over a 45-month window and were used as the covariates, together with other profile data.

For the average revenue analysis, a seemingly unrelated regression (SUR) model (i.e. structured equations model) was built to forecast average revenue per customer (ARC) based on the frequency, duration and amount of purchase/service. The dependent variables were accumulated over a 45-month period and the independent variables are account details and/or customer profile data (including tenure).

4. Data

Transactional data were derived over the most-recent (at the time of study) 45-month period. A total of 2,691 customers were included in the sample, of which about 51% are active. Majority were personal account-type (85%), and 12% were corporate and the remaining are retailers. Ninety-four percent are regular customers, mostly females (at 55%) and at 30-40 years old (33%) or 40-50 years old (29%). The distribution by joining year (Year 0 as most recent at the time of study, Year -1 is one-year back, Year -2 is two-years back, and so on) is given in Table 1.

Table 1. Distribution of Sample by Year of Joining

Year	Frequency	Percentage
-11	42	1.32%
-10	90	2.83%
-9	89	2.79%
-8	135	4.24%
-7	92	2.89%
-6	272	8.54%
-5	337	10.58%
-4	394	12.37%
-3	446	14.00%
-2	728	22.86%
-1	404	12.68%
0	156	4.90%

The tenure data is composed of two types – the completed tenure (i.e. historic), and the truncated tenure (i.e. active). Note that for the second approach on tenure analysis, non-censored data are those customers whose lifetime ended and whose start (or joining) dates fall within the observed period. Left-censored data happens when a historic customer's joining date is prior to (or to the left of) the observation period. Right-censored data happens when an active customer's joining date is within the observation period. Interval-censoring happens for active customers whose commencement date is prior to the observation period. Thus, completed tenure customers have either non- or left- censored data, while truncated tenure customers have either right- or interval- censored data.

5. Results and Discussion

Tenure Analysis via Hazard Function Model

Survival probability function (i.e. the expected lifetime distribution) across all data (and all groups) follows a Weibull distribution. Over-all EL at the 50th percentile (EL50) is at 75 months, that is, for a given customer, there is a 50% chance of survival come his 75th month of engagement with the company. The plot of the survival function is found in Figure 1.

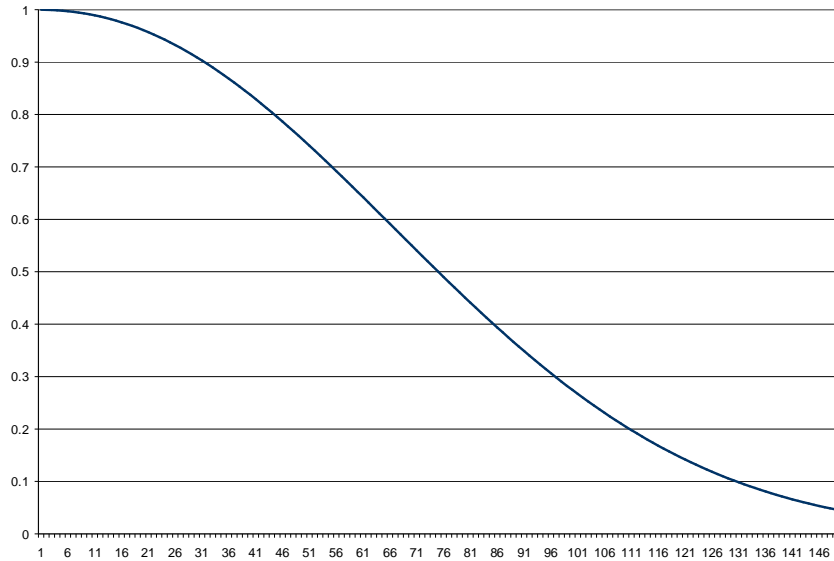


Figure 1. Over-all Survival Model (i.e. EL model)

Across age groups, those aged less than 30 years old had the lowest probability of survival (i.e. staying with the company) with EL50 only at 5.5 years, while those aged over 50 years old had the highest probability of survival with EL50 at 6.5 years. Those within the 30-39 years old and the 40-50 years old were relatively at mid-level, with EL50 at 6 years. The plot of the survival functions are in Figure 2.

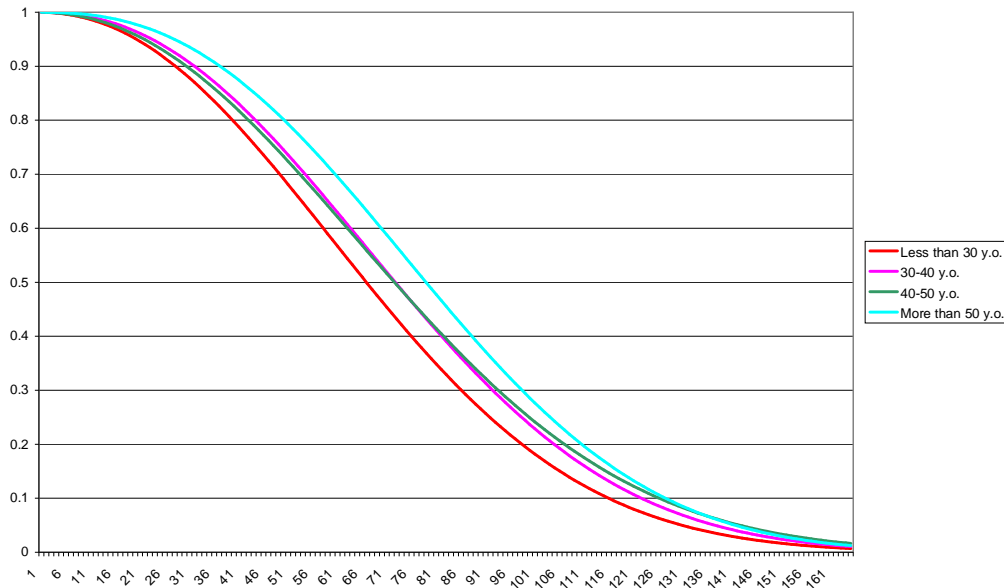


Figure 2. EL models across Age Groups

Across account type, corporate accounts had the highest likelihood of survival within the first 10 years, with EL50 at 7 years. Retailers had the lowest likelihood during the first 7 years but had the highest likelihood after 10 years, and with EL50 only at 5 years and 5 months. Personal accounts are at mid-level during the first 7 years and become lowest thereafter, with EL50 at 6 years and 1 month. The distribution functions are in Figure 3.

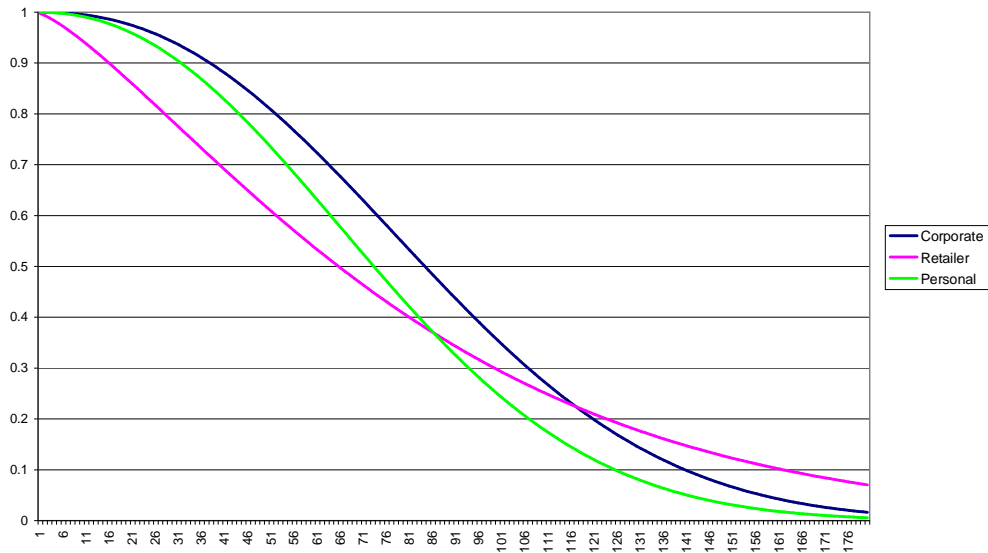


Figure 3. EL models across Account Type

Across gender, the hazard functions are comparable but with that of males relatively higher. Females' EL50 is at 6 years while that of males is at 6.5 years. The plots are in Figure 4.

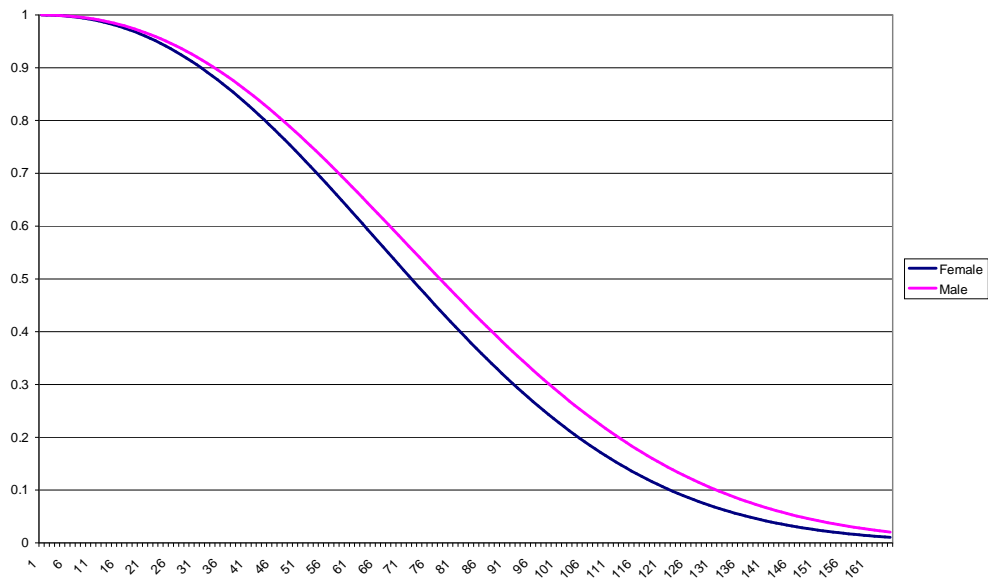


Figure 4. EL models across Gender

Differences in market segments are very much evident, with very high discrepancy between regular and premium customers (see Figure 5). EL50 for regular customers is nearly 6 years, while that of premium customers is more than 10 years.

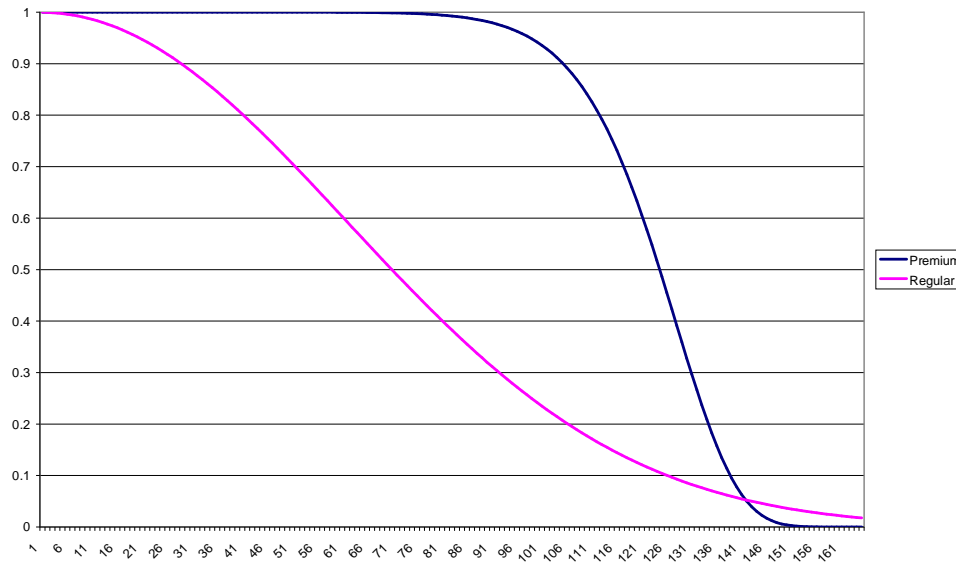


Figure 5. EL models across Market Segment

Tenure Analysis via Generalized Regression Model

Different models were considered with the covariates present in the analyses. Payment modes were also considered in modeling – whether customers pay in full or partial, on-time or late, or at a specific mode (cash, cheque, etc.). Unsegmented (i.e. all data included) and segmented (i.e. for each pre-identified groups – account type, market segment) models were generated as a pre-process for identifying possible indicators for the final EOL model. The final EOL model had the following significant covariates – indicator if premium, indicator for account type, tenure (at the beginning of the accumulation period), indicator if account is over 2 years, and usage details (frequency, duration and amount of specific services). The EOLs (average additional tenure/remaining years with the company) of the sample data across different profiles are summarized in the following table (but note however that these were computed based on average values on the transaction details).

Table 2. Average EOL by Profile

	Profile	Average EOL (in Years)
Over-all	<i>(all data)</i>	2.85
Account Type	Corporate	2.22
	Retailers	1.46
	Personal	3.05
Age Group	Less than 30 y.o.	3.68
	30-40 y.o.	3.04
	40-50 y.o.	3.00
	Over 50 y.o.	2.83
Gender	Female	3.13
	Male	2.90

ARC Analysis via SUR model

To model ARC, the independent variables considered were indicators on customer's age, gender, account type, activity (historic or active) and tenure, together with number of supplementary accounts. Note that three different sets of dependents were considered (on frequency, on duration, and on amounts, across different services/products) via structural equations (with dependent variables as exogenous within a set), and thus ARC can be predicted either through regularity of usage/purchase, extent of use, or outright cost.

Results suggest that effects of the independent variables vary across each of the dependent variables, and as a summary, the table below presents average frequencies, durations and amounts at zero-valued indicator variables and at-minimum quantitative variables.

Table 3. Monthly Average ARCs per Product

	Frequency	Duration (Hours of Use)	Amount
<i>FreeService1</i>	7.48	14.39	
Product1A	152.23	5.74	1,574.58
Product1B	~0	~0	~0
Product1C	1.22	0.02	108.33
Product1D	0.43	0.0008	2.32
Product2A	417.75		283.02
Product2B	194.52		125.74
Product2C	9.92		90.18
Product2D	3.50		62.00
Product3	96.53	15.18	135.19

Integration of Tenure and ARC models

In the absence of transactional data, the hazard function models can be used in scoring for CLV, with customers having high likelihood to stay to be given higher CLV score and better incentives. The CLV scoring however becomes more dynamic and/or efficient when referenced to static data, in this case via the different transactional and profile data. Prioritization scheme can then be implemented on a timely basis. High-valued loyal customers are those perceived to stay longer with the company and are expected to have high engagement (in terms of usage/purchase of services/products, and thus higher profitability). Devoted but unexcited customers are those perceived to have high tenure but have low engagement, thus up-sell and/or cross-sell efforts may be employed. Urgent targets are those perceived to terminate soon but have high engagement, thus the company must come up retention campaigns.

Single scoring, on the other hand, entails more specific company directives. With the current tenure and ARC models, the immediate next step is to identify appropriate "weights" aligned with the company objectives. If profitability is of top priority, then weights may be skewed towards high predicted ARC. Otherwise, if retention is priority, then weights may be assigned heavily to those with high likelihood to stay with the company. If profitability and retention are of equal importance, then weights may be equally distributed.

6. Conclusion

For companies whose marketing strategy is centered on what the customer needs or expects, or whose business objective evolves around customer experience, CLV modeling remains a vital tool in meeting such directives. Though CLV models take on different approaches, different data mining techniques are proven useful in understanding deeper the behavior of the customers and at the same time assess the value of customers on a timely basis.

References:

- Bauer, H., Hammerschmidt, M., and Braehler, M., 2003, The Customer Lifetime Value Concept and its Contribution to Corporate Valuation, *Yearbook of Marketing and Consumer Research*, 1: 47-67.
- Berger, P., and Nasr, N., 1998, Customer Lifetime Value: Marketing Models and Applications, *Journal of Interactive Marketing*, 12(1): 17-30.
- Berger, P., Eechambadi, N., George, M., Lehmanm. D., Rizley, R., and Venkatesan, R., 2006, From Customer Lifetime Value to Shareholder Value: Theory, Empirical Evidence, and Issues for Future Research, *Journal of Service Research*, 9(2):156-167.
- Dwyer, F., 1989, Customer Lifetime Valuation to Support Marketing Decision Making, *Journal of Direct Marketing*, 3(4):8-15.
- Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., and Sriram, S., 2006, Modeling Customer Lifetime Value, *Journal of Service Research*, 9(2): 139-155.
- Lee, J., Lee, J., and Feick, L., 2006, Incorporating Word-Of-Mouth Effects in Estimating Customer Lifetime Value, *Database Marketing and Customer Strategy Management*, 14(1): 29-39.
- Malthouse, E., and Blattberg, R., 2005, Can We Predict Customer Lifetime Value?, *Journal of Interactive Marketing*, 19(1):2-16.
- Reinartz, W., and Kumar, V., 2003, The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration, *Journal of Marketing*, 67:77-99.
- Rosset, S., Neumann, E., Eick, U., Vatnik, N., 2003, Customer Lifetime Value Models for Decision Support, *Data Mining and Knowledge Discovery*, 7:321-339.
- Ryals, L. and Knox, S., 2005, Measuring Risk-Adjusted Customer Lifetime Value and Its Impact on Relationship Marketing Strategies and Shareholder Value, *European Journal of Marketing*, 39:456-472.
- Venkatesan, R., and Kumar, V., 2004, A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy, *Journal of Marketing*, 68: 106-125.

