Credit Card Lifetime Value Modeling: The Case of a Large US Financial Service Corporation

Liangdong Huang, The University of the Incarnate Word

Adam Guerrero, The University of the Incarnate Word

ABSTRACT

The purpose of this research is to build multivariable statistical models used to examine the drivers of credit card customer lifetime value (CLV). To accomplish this goal, a sample of 500,000 credit card customers were randomly sampled from a large US bank. The dependent variable is a customer-specific profitability metric used to approximate credit card CLV. Three categories of determinants were used to explain the variation of credit card CLV including customer demographic characteristics, company relationship variables, and variables that measure credit card account behavior. A one-year observation period was used to collect data on the dependent and independent variables. A general linear regression model was initially constructed for the entire sample before the estimation of separate models for customer segments. Statistically significant findings were obtained for all models estimated, with customer segmentation models outperforming the results from the pooled regression.

INTRODUCTION TO STUDY

Customer lifetime value (CLV) evolved as a relevant concept in the customer relationship management literature and has achieved industry-wide implementation since 1990. CLV measures the long-term profitability for both new and existing customers. The CLV literature has focused on CLV concepts and applications, theoretical frameworks used to construct CLV models, and the empirical estimation of CLV Models based on profitability data. Although there are a few peer-reviewed papers that estimate CLV models based on actual customer and transaction data, most of them forecast short- versus long-term profitability using 1 or 2 years of data. Since Customer lifetime value (CLV) is a forward-looking, long-term profitability metric, modeling it requires the use of many years of data. Further, many papers published in the CLV literature provide a conceptual mathematical framework for modeling CLV using illustrative mock-up examples, but few papers empirically model CLV using real-world data. The aim of this research is to empirically investigate the determinants of credit card CLV models built using long-term profitability data from a large US retail bank.

In this study, a series of multivariate regression models are estimated to evaluate the factors that impact credit card customer lifetime values, or the long-term profitability of credit card customers at a large financial service corporation. The dependent variable is credit card customer lifetime value, measured as the net present value of the annual profits over an eight-year time span. The independent variables include customer demographic information such as age, gender, education, marital status, income, and geographic location; customer risk assessed using credit scores; credit card account data such asaccount status, total purchases, revolving balances, recency of usage, and frequency of usage; and customer relationship metrics including company tenure, product tenure, total number of other products, and other major product ownership. Again, all data are sampled from a retail bank located in the United States.

Background of the Study

Customer lifetime value (CLV) is defined as the present value of a customer's future profitability relative to cost cash flows (Berger & Nasr, 1989; Chang, 2016; Gupta et al., 2006; Ekinci, 2014). Customer lifetime value models enable companies to identify more profitable customers and to optimize resource utilization for customer acquisition and retention initiatives (Aeron et al., 2008; Berger & Nasr, 1998; Chang, 2016; Dwyer, 1989; Ekinci, 2014; Gupta et al., 2008; Marmol, 2021; Memarpour, 2019).

CLV models incorporate several key components including customer and product acquisition probabilities, revenue and cost estimates, the retention rate, and the discount rate. In estimating CLV models, most researchers include only some CLV factors while applying assumptions to others, however Kumar and Reinartz (2016) recommend modeling all components together. Sincecustomer lifetime estimates vary by industry, for example, due to varying contractual requirements related to the consumption of products, the amount of data used by researchers will also vary. Interestingly, most CLV models include less than five years of data (Donkers et al., 2007; Ekinci, 2014) even through actual customer lifetimes could endure for significantly longer periods of time.

A credit card is a retail banking product adopted by customers as both a payment instrument and a short-term lending vehicle. After customers make purchases, they can pay off credit card statement balances or pay a minimum amount and borrow the remaining balance, which will incur interest charges. Banks profit from both interest income and non-interest income but also incur losses when customers do not repay their remaining balances. Despite the implementation of customer acquisition and account management strategies at the customer level, most banks do not generate sufficient customer level profitability data, and therefore do not construct and operationalize customer level CLV models (Berger & Nasr, 1998; Finlay, 2008; Osipenko, 2018).

LITERATURE REVIEW

Areon et al. (2008) developed a conceptual framework to explain the relationship between organizational decision-making, revenue performance, and customer borrowing and payment decisions. This conceptual model primarily focused on customer state transitions from one period to the next but did not explain the factors that impact customer behavior. Kumar and Reinartz (2016) developed a framework that examines the impact that firm product and service offerings, price decisions, and marketing programs have on customer perceptions and behaviors, as well as the ensuing relationship between customer behavior and firm profitability. Hence Kumar and Reinartz (2016) developed an organizing framework that examines the relationship between customer perceived value with the value of a firm. Gao et al. (2020) developed an even more elaborate conceptual framework of customer equity. In this model, customer profitability is influenced by the quality of the customer experience, which is influenced by key aspects of the customer experience including value equity, brand equity, and relationship equity, all of which can be affected by various social influences including family, friends, and social media.

The Concept of Customer Lifetime Value

The topic of customer lifetime value evolved from research that examines many of the customer management practices. A key concept underpinning the customer management practices literature is the idea that customers have distinct needs, preferences, and behaviors that, in turn, contribute to the generation of a diverse set of organizational values (Ekinci, 2014). The customer management practice literature originated in the early twentieth century at the time when companies began to shift their attention from inactive customers by removing them from mailing list campaigns to more reliable customers associated with higher profit margins (Oblander, et al., 2020). The first customer lifetime value applications started to appearin late 1960s when companies such as Reader's Digest Magazine used ZIP codes to target more profitable customers. Shortly thereafter, organizations began to use customer relationship management (CRM) databases, which enabled companies like American Airlines and American Express to implement customer reward programs, encouraging repeat purchases of products and/or services (Oblander, et al., 2020). The Customer Lifetime Value (CLV) concept was formalized in the 1990s by Peppers and Rogers with the introduction of a new marketing paradigm entitled "share of customer," rather than just market share, and then the "loyalty effect" by Reichheld (insert date). The loyalty effect is the idea that companies can generate greater profit from customers by maintaining their loyalty for longer durations, increasing product sales as a result (Ekinci, 2014; Oblander, et al., 2020).

Since 1990, CLV research can be categorized into three mean areas of inquiry. The first category explores different ways to measure, or calculate customer lifetime value using a variety of illustrative examples (Aeron et al., 2008; Berger & Nasr, 1989; Calciu, 2009; Dwyer, 1989; Gupta et al., 2006; Pfeifer & Carraway, 2000). The second domain of inquiry focuses on how to employ information pertaining to CLV analyses to enhance business decision-making in the areas of marketing expense allocation (Memarpour et al., 2019; Venkatesan & Kumar, 2004) and sales growth (Caseas-Arce et al., 2017). The third category explores a variety of techniques used to model CLV (Brrios & Lansangan, 2012; Chang & Ijose, 2016; Costa et al., 2018; Däs et al., 2017; Dunkers et al., 2007; Ekinci et al., 2014; Estrell-Ramon et al., 2017; Haenlein et al., 2007; Jasek et al., 2019; Li et al., 2012; Rezaei et al., 2022; Yoo et al., 2020).

CLV Definitions

Unlike backward looking short-term profitability measurements based on historical data, customer lifetime value **(CLV)** measures long-term profitability, a metric that is forward-looking by nature because it is mainly concerned with future earnings. Customer lifetime value (CLV) can be solved for as the net present value of future profits, or cash flows, expected to be received from an individual customer (Berger & Nasr, 1998; Calciu, 2009; Chang & Ijose, 2016; Costa et al., 2018; Dwyer, 1989; Ekinci, 2014; Gupta et al., 2008; Kumar & Reinartz, 2016). CLV can be solved for using simple time value of money formulas. A proxy for CLV is net present value (NPV) solved for as:

$$\mathbf{CLV} = \sum_{t=0}^{T} \frac{P_t}{(1+i)^t} - \mathrm{AC} \tag{1}$$

where

t is the time period,
T is the time horizon to calculate CLV,
P_t is the profit at time t when the customer is still active,
i is the discount rate or cost of capital of the company, and
AC is the acquisition cost of the customer

Formula (1) can be expanded to include profit calculation components, such as the sales price, the cost of goods/services sold, and the retention rate, which can be solved for as the portion of profits retained by the firm versus being paid out in the form of dividends. If we assume the profit margin and retention rate remain constant over time and that customer lifetime is infinite, then the CLV calculation can be simplified to the margin multiplied by $R/(1+i-R)$ (Gupta et al., 2006).

$$\mathbf{CLV} = \sum_{t=0}^{T} \frac{(P_t - C_t) * R_t}{(1+i)^t} - \mathrm{AC} = \mathrm{Margin} * \frac{R}{(1+i-R)} \tag{2}$$

where

t is the period
P_t is the sales price at t when the customer is still active
C_t is the cost of product/services at t when the customer is still active

R_t is the retention probability at time t
i is the discount rate measured by the cost of capital
AC is the acquisition cost of the customer
T is the time horizon to estimate the CLV

Formula (1) and (2) can be expanded to account for the fact that acquired customers are often retained to consume more than one product over their lifetime (Chang & Ijose, 2016).

$$\text{CLV} = \sum_{t=0}^{T} \sum_{j=1}^{J} \frac{A_{jt} * (P_{jt} - C_{jt}) * R_{jt}}{(1+i)^t} - AC \tag{3}$$

where

t is the time period
j is the j^{th} product a customer will acquire
A_{jt} is the acquisition probability at time t. A_{jt} will be 100% after the product is acquired.
P_{jt} is the sales price for product j at time t
C_{it} is the cost of product/services for product j at time t
R_{jt} is the retention probability for j^{th} product at time t
i is the discount rate or cost of capital of the company
AC is the acquisition cost of the customer
T is the time horizon to estimate the CLV

CLV formulae can be expanded to include customer referral value (CRV, see formula 4), which is solved for as the total CLV of individuals recommended to the firm by the existing customer. In an analysis of a sample of retirement plan customers, Costa et al. (2018) found a U-shape relationship between an existing customer's CLV and the total lifetime value of referrals (i.e., CRV). Further, Costa et al. (2018) clustered, or segmented, customers into three groups, "Catalysts," "Neutrals," and "Stars," all of which had varying lifetime values. The first segment referred to as "Catalysts" had low CLVs due to details surrounding their retire plans. However, since "" Catalysts" had high recommendation power, there were able to sustain high CRV values. The second group, "Neutrals", did not make significant recommendations, and therefore had medium CLVs. The third group, "Stars", had higher CLVs associated with a higher level of recommendations. Finally, customers with high satisfaction scores are more likely to recommend firm products and services to friends and family. The willingness of a customer to refer a product to friends and family can be measured by the net promoter score (NPS), which was developed by Reichheld at Bain & Company in 2003.

$$\mathrm{CRV_i} = \sum_{j=1}^{J} CLV\ of\ Referral_j \tag{4}$$

where

i is the i^{th} existing customer who made referrals
j is the j^{th} referral who becomes new customer

Given the increasing importance of social media for marketers, Däs et al. (2017) conducted a study of customer lifetime network value (CLNV), that expanded Costa's word-of-mouth research to include an internet-based social network recommendation effect. An individual can leverage his/her social network to increase the overall awareness of an organization's goods and/or services by sharing important information regarding the customer experience, such as the quality of customer service, sales service, price, etc., all of which may contribute to a higher customer acquisition rate. An individual's customer lifetime network value (CLNV) includes his/her own customer lifetime value (CLV) plus his/her customer referral value (CRV), as described in formula 5 below. CRV can be positive, zero, or negative, depending on the value of the sum of the present values of customer contributions in his/her network. The network referral effect can be small or even zero if customers make purchase decisions independently.

$$CLNV_i = CLV_i + CRV_i \tag{5}$$

where

i is the i^{th} existing customer

CLV can be measured at the level of an individual customer, or it can be aggregated by customer segment or cohort (Kumar & Reinartz, 2016). CLV can even be aggregated to solve for an organization's Customer Equity (CE), measured as the total value of CLVs for currently existing and future customers (Gupta et al., 2008; Kumar & Reinartz, 2016; Oblander et al., 2020). The drivers of CLV include the acquisition rate, retention rate, cross-sell rate (Gupta et al., 2006), and indicators of product profitability. Marketing programs directly impact the response rate, retention rate, and the product expansion rate. Product profitability is influenced by price and the nature and quality of customer interactions with the firm.

CLV Applications

Data analyses centered on CLV make it possible for firms to target and retain more profitable customers. Before organizations practiced CLV analysis, marketing departments focused on top-of-the-funnel sales metrics including customer awareness, interest, consideration, intent, evaluation, and buying behavior. On the other hand, CLV analyses allow marketers to better assess the long-term financial impact of customers and/or customer segments via bottom-of-the-funnel sales metrics, resulting in a more efficient allocation of scarce marketing resources (Ekinci et al., 2014).

Customer lifetime value analytics can be used to assess the profitability of prospects, new customers, and existing customers. CLV is applicable to multiple industries, sectors, market structures, and business transactions including business-to-business, business-to-consumer, consumer-to-business, and consumer-to-consumer purchases. (AboElHamd et al., 2020). A majority of CLV components - for example, purchase frequency, the retention rate, and whether customers buy additional products and services – are influenced by a company's customer relationship strategy (Costa et al., 2018).

Customer lifetime value applications can improve the return on acquisition investment by identifying more profitable customers (AboElHamd et al., 2020; Ekinci, 2014; Gupta et al., 2008), directing mailing campaigns to more profitable customers with high response rates (Oblander, et al., 2020), efficiently allocating marketing resources by media type, (Pfreifer & Carraway, 2000), prioritizing the amount of time spent on different customers classes(Casas-Arce, 2017), and setting a ceiling on acquisition spending (Berger & Nasr, 1998; Dwyer, 1989).

Traditionally, four types of variables have been used to segment markets: customers demographics, geographic information, psychographic data, and characteristics of consumer behavior. When organizations evaluate CLV by customer segment, they are able to identify the most profitable customers, or customer groups (Ekinci, 2014). They are also able to devise strategies to enhance profits for lower margin customer groups. These tactics can be further improved upon when organizations supplement CLV information by segment with customer referral value by segment. e For “influenced champions and classic champions” customer groups who had high customer lifetime values yet low or negative customer referral values, firms can provide referral awards to encourage customers to recommend products to their friends and relatives (Däs et al., 2017).

Two additional CLV applications are related to customer retention and customer equity. Research has shown that it is more expensive to acquire a new customer than to retain an existing customer (insert citation here using the appropriate article below). Investigating how CLV varies for currently existing customer groups makes it possible for organizations to apply different servicing strategies so that they can devote resources to retaining customers worth keeping (Berger & Nasr, 1998; Chang &Ijose, 2016; Costa et al., 2018; Ekinci, 2014; Kumar & Reinartz, 2016). Furthermore, customer lifetime values categorized by customer segment can be aggregated to calculate the value of an organization's customer equity, which can then be used to evaluate overall firm competitiveness and merger and acquisition decisions (Däs et al., 2017; Gupta et al., 2006; Venkatesan and Kumar 2004).

Modeling Customer Lifetime Value

While the concept of CLV is straightforward, the construction of empirical CLV models can be complicated due to multiple profit and cost components. CLV models also vary by product or service type, further contributing to their complexity. The literature dedicated to modeling CLV includes both mathematical calculations using illustrative data (Aeron et al., 2008; Berger & Nasr, 1989; Calciu, 2009; Dwyer, 1989; Gupta et al., 2006; Pfeifer & Carraway, 2000) and statistical/machine learning models using customer relationship management data (Brrios & Lansangan, 2012; Chang & Ijose, 2016; Costa et al., 2018; Däs et al., 2017; Dunkers et al., 2007; Ekinci et al., 2014; Estrell-Ramon et al., 2017; Haenlein et al., 2007; Jasek et al., 2019; Li et al., 2012; Rezaei et al., 2022; Yoo et al., 2020).

In addition to considering customer type, CLV models must be adjusted to consider the type of good or service under analysis. As an example, Dwyer (1989) discussed two types of customers: "lost-for-good" customers" and "always-a-share" customers. An example of the "lost-for-good" customer type are mobile phone service contract customers characterized by making long-term firm-specific commitments to avoid switching costs, also known as breakout contract penalties. Since mobile phone bill pay/revenue is relatively stable, or fixed, over time (mobile phone contracts are set at relatively stable monthly rates), there is little need to construct complicated statistical models. Further, CLV models for "lost-for-good" customers tend to put more emphasis on the customer retention rate, or attrition rate, relative to CLV models for other types of goods and customers (Dwyer, 1989). Credit card customers are an example of the "always-a-share" customer group since they generally own multiple credit cards issued by different organizations. For a specific credit card issuer, the wallet-share from a cardholder can range from 0 to 100%. Since credit card customers can be active, inactive, or even reactive, the wallet share from a cardholder also varies over time, leading to unstable revenue streams. Thus, CLV models for "always-a-share" customers must take into consideration both variable revenue streams over time and the customer retention rate.

Researchers can create one model using CLV as the dependent variable or they can construct more than one model using as dependent variables various CLV components including but not limited to sales, profit margins, various cost components, the customer acquisition rate, and the customer retention rate. Furthermore, researchers predict CLV components using many different types of independent variables, some of which are commonplace in the literature, and others much more unique. Kumar and Reinartz (2016) argue that it is more accurate to model CLV and its components together.Costa et al. (2018) and Däs et al (2017) extended CLV models to include customer referral value as measured by word of mouth or indicators that measure the impact of customer social networks. Although models that estimate CLV are supposed to measure the lifetime benefits associated with firms maintaining successful long-term customer relationships, most CLV research includes less than five years of data on profits (Donkers et al., 2007; Ekinci, 2014). However, it might be appropriate to model CLV using five years of data or less because the long-term profits associated with maintaining customer relationships becomes less and less material over time – e.g., the net present value of profit expected to be received in six years is less than half of the original amount with a discount rate equal to approximately twelve percent. Since it is often difficult for researchers to collect more than five years of customer profitability data, models of CLV are often illustrated using a potentially oversimplified mathematical demonstration

(Dwyer, 1989; Pfiefer and Carraway, 2000). More research needs to be done to determine the optimal amount of time periods to include in CLV analyses, which might vary by study for good reason.

CLV Model Dependent Variables

Different researchers use different dependent variables to construct CLV models. Also, it is not uncommon for researchers to model various CLV components before using those results to calculate CLV. For example, most researchers developed simplified CLV models that predict future revenue streams given assumptions surrounding various CLV components including margins and costs (Kumar & Reinartz, 2016). Assumed values are often based on historical averages or known statistical distributions highlighted in the literature, such as the generalized gama distribution (Venkatesan & Kumar, 2004). CLV models CLV components include short-term annual profits (Donkers, 2007; Ekinci et al., 2014; Jasek et al., 2019), revenues (Barrios & Lansangan, 2012; Costa et al., 2018;), sales proceeds from specific goods and services (Chang & Ijose, 2016; Yoo et al., 2020), customer account balances (Li et al., 2012), the number of transactions (Li et al., 2012), contribution margins (Donkers, 2007; Estrella-Ramon et al., 2017; Haenlein et al., 2007), expenses (Chang & Ijose, 2016), the customer retention rate (Chang & Ijose, 2016; Costa et al., 2018; Donkers, 2007), purchase frequency (Estrella-Ramon et al., 2017; Yoo et al., 2020), customer lifetimes, or durations (Barrios & Lansangan, 2012; Donkers, 2007), and the discount rate (Estrella-Ramon et al., 2017).

CLV Model Independent Variables

Researchers used a variety of independent variables in CLV models centered on predicting CLV or its components including profitability measures, revenues, costs, the customer retention rate, and so on. Independent variables used to predict CLV can be categorized into three groups:

customer-level variables, firm-specific variables, and variables related to the external environment. Customer-specific variables include demographic information, such as age (Barrios & Lansangan, 2012; Chang & Ijose, 2016; Costa et al., 2018; Donkers et al., 2007; Estrella-Ramon et al., 2017; Haenlein et al., 2007), gender (Barrios & Lansangan, 2012; Costa et al., 2018; Estrella-Ramon et al., 2017; Haenlein et al., 2007), marital status (Chang & Ijose, 2016; Costa et al., 2018; Haenlein et al., 2007), family size (Chang & Ijose, 2016), education (Costa et al., 2018; Haenlein et al., 2007), income (Chang & Ijose, 2016; Estrella-Ramon et al., 2017), occupation (Chang & Ijose, 2016), geographic location (Chang & Ijose, 2016; Costa et al., 2018), company tenure, product ownership (Barrios & Lansangan, 2012; Costa et al., 2018; Donkers et al., 2007; Ekinci et al., 2014; Estrella-Ramon & Sanchez-Perez, 2017; Haenlein et al., 2007), product usage, purchase quantity, cross-buying behavior, (Chang & Ijose, 2016; Das et al., 2017; Ekinci et al., 2014; Estrella-Ramon & Sanchez-Perez, 2017) and recency of purchase (Costa et al., 2018; Gupta et al., 2006). Rezaei et al. (2022) included omnichannel (digital and physical) usage behavior while Estrella-Ramon & Sanchez-Perez (2017) included the adoption of online banking services as explanatory variables. An additional regressor found in the literature is customer perceived value, defined as how customers view the overall benefits, costs, and undesired consequences of consuming a good or service, a concept for which the level of customer satisfaction is often used as a proxy (Ho et al., 2006; Kumar & Reinartz, 2016). However, since CLV is calculated for each customer while customer satisfaction surveys are distributed to a small percentage of customers, the level of customer satisfaction results in restricted data sets that limit its use as a building block in the construction of CLV models.

Variables measured at the firm level include the price of the good or service, the marketing budget, customer loyalty programs (Donkers et al., 2007; Yoo et al., 2020) and product bundling discounts (Donkers et al., 2007). Variables related to the external environment include macroeconomic indicators (Gupta et al., 2006; Jasek et al., 2019; Yoo et al., 2020), sociocultural factors (Jasek et al., 2019), periodic seasonal fluctuations in sales (Yoo et al., 2020), and factors that measure the level of industry competition (Gupta et al., 2006).

CLV Model Type

Researchers have used a variety of methodological approaches to model CLV and its components. Examples include recency, frequency, and monetary (RFM) analysis, Pareto/negative binomial distribution (NBD) analysis, Markov chain model (MCM) estimation, ordinary least squares regression analysis, Bayesian hierarchical modeling, survival analysis, artificial neural networks (ANN), and classification and regression trees (CART) (Ekinci et al., 2014; Gupta et al., 2006; Kumar & Reinartz, 2016).

Recency, frequency, and monetary (RFM) analysis uses historical data pertaining to the recency, frequency, and monetary value of customer purchases to segment customers. RFM models represent one of many direct marketing methods focused on improving the response rate of marketing campaigns targeted at customer segments constructed using RFM purchase data.
RFM analysis was first introduced by Hugh in 1994 to group customers based on the recency, frequency, and monetary total of customer transactions. Mathematically, RFM models tend to be more descriptive than inferential, using recency, frequency, and monetary total data to assign customers to groups. Chang and Tsay expanded RFM analysis to include the length of time an individual remains a customer in addition to recency, frequency, and the monetary value of purchases, known as LRFM analysis (Alizadeh Zoeram & Karimi Mazidi, 2018). By including past purchasing behaviors in addition to socioeconomic information, RFM models outperformed traditional market response models, which relied too heavily on demographic data to classify

prospects or existing customers who respond to marketing initiatives. One limitation of RFM analysis is that it results in a short forecast window, reducing its ability to predict future customer behavior (Gupta et al., 2006). A limitation of CLV models constructed using RFM transaction data is that CLV predictions are too heavily influenced by past marketing campaigns versus future initiatives (Gupta et al., 2006). RFM models are not used to calculate CLV directly, rather they are used to construct categorical indicators that are then used to forecast CLV (Gupta et al., 2006). As an example, Rezaei et al. (2022) conducted a survey of 330 managers to classify customers based on six RFM variables, which were then used to examine CLV. RFM analysis is most often used as an exploratory technique to segment customers into groups such as potential customers, core customers, new customers, lost customers, and resource-heavy customers (Alizadeh Zoeram & Karimi Mazidi, 2018).

Pareto/Negative Binomial Distribution (NBD) models are used to find the lifetime value of customers in non-contractual settings, where a customer can make purchases at any time (Gupta et al., 2006) and customer lifetimes are uncertain (Li et al., 2012). The Pareto/Negative Binomial Distribution extends RFM analysis by using an orders table that contains information pertaining to the recency, frequency, and monetary value of customer purchases to model two distributions commonly found in the CLV literature, customer lifetimes and transactions. The Pareto distribution is used to model customer churn, also known as customer as customer attrition, or dropout, and the negative binomial distribution is used to model customer transactions, or purchase frequency. Predicted lifetime values can be obtained by simply multiplying the expected lifetime from the Pareto distribution by the expected purchases from then negative binomial distribution, all multiplied by the average purchase amount. In an assessment of customer saving accounts, Li et al. (2012) used an NBD model to forecast transaction time and the Gamma-Gamma distribution, commonly used in Bayesian statistical analyses, to forecast monthly average account balances. In a comparative analysis of eleven Pareto/NBD model variations, Jasek et al. (2019) analyzed 2.3 million online shoppers completing approximately 3.8 million transactions. Interestingly, no single model outperformed all other variations based on commonly used model selection criteria. Finally, NBD models can be used to evaluate the lifetime values of credit card customers since credit card use represents a non-contractual relationship given customers have a strong repayments status characterized by keeping their balances below their credit limits.

In the context of modeling CLV, Markov Chain analysis is used to estimate the transition probabilities of customers moving from one behavioral state to the next, also known as customer state migrations over time, with states representing stages in the purchasing process or segments that customers can transition into and out of during their lifetimes, affecting their lifetime values (Aeron et al., 2008; Dwyer, 1989; Haelein et al., 2007; Pfeifer & Carraway, 2000). Applications found in the literature include Dwyer (1989), who illustrated how to calculate CLV using state migration probabilities for four time periods (T_4) after customers made a first purchase (T_0). Pfiefer and Carraway (2000) using Markov Chain analysis to examine state migrations probabilities for customer segments including new prospects, existing customers, and former customers. Haelein et al. (2007) used a Markov chain to model the purchasing behavior changes of customers, specifically how they transition into and out of known customer segments. Aeron et al. (2008) examined state transition probabilities calculated based on historical averages for the following customer segments: acquired, inactive, transact, revolve, delinquent, default and attrite. After segment transition probabilities were solved for, the state migration matrix was used to simulate CLV for 100 months. A limitation of the use of Markov Chain analysis to solve for state transition probabilities based on historical averages for all customers combined is that estimated state probabilities might not hold at the customer segmentation level (Ekinci et al., 2014). Finally, the Markov Chain model is often used in conjunction with additional analytical techniques used to estimate key components of CLV models including churn probabilities, customer lifetime durations, profitability metrics, etc. Finally, a wide variety of multivariate regression techniques are used by researchers to predict CLV and its component parts.

METHODOLOGY

In this study, a random sample of customer data was used to construct credit card customer CLV models as a function of customer relationship variables, customer behavior variables, and customer demographic criteria. Approximately 500,000 bank customers were used to estimate a pooled model that was then reexamined for three separate customer segments previously defined by the bank as inactive customers, revolver customers, and transactor customers. As Table 2 shows, the study leveraged a cohort design comprised of multiple years of annual profits, which were then used to solve for customer lifetime values. More specifically, the present value of cash expected to be received from customers between 2011 to 2018 was used as the basis for the dependent variable. A twelve-month observation period from January 2010 to December 2010 was used to collect data on independent variables including the customer utilization rate, maximum delinquency status, recency, total transactions, total spending, etc. Certain characteristics were measured using data from the most recent month, such as credit scores, whereas other independent variables were measured based on a 12-month history. It Although customer origination credit scores might exceed the 12-month observation window, all independent variables precede measurements taken on the dependent variable. It should be noted that a small percentage of customers were excluded from the analysis due to voluntary attrition or involuntary charge-offs. Data were merged at the customer level for samples with more than one credit card. Independent variables that have either a non-monotonic or a non-linear relationship with the dependent variable were recoded into a series of categorical variables. Table 2 provides a brief overview of the variables used in this study, their measurement, and descriptive statistics. Customer profitability metrics used in solving for customer lifetime values are omitted to key confidentiality. Further, 95.5% of customers are current, 3.8% are 1 to 29 days past due, and less than 1% for 30 days past due. Finally, 11% of the customers are in active, 50% of the customers are revolvers, and 39% of the customers are transactors.

Table 2 Descriptive Statistics

Variable	Description	Mean	Variable
Customer demographic variables			
Age	Age measured in years	48	17
Income	Measured in dollars	51139	47139
OFICO	Origination credit score	614	277
Current FICO	Refreshed FICO score at observation time	729	128
Customer relationship variables			
Company tenure	Months since customer joined bank	221	164
Card tenure	Months since customer acquired credit card	128	96
Total products	Total bank products customers own	2.10	1.52
Checking	Flag of owning checking product (1=Yes)	0.34	0.48
Saving	Flag of owning saving product (1=Yes)	0.29	0.45
CD	Flag of owning time deposit (1=Yes)	0.03	0.16
Consumer loan	Flag of owning consumer loan (1=Yes)	0.15	0.35
Mortgage	Flag of owning mortgage loan (1=Yes)	0.06	0.25
Customer behavior variables			
Recency	Most recent month had transaction	1.41	1.88
Frequency	Number of transactions in last 12 months	108	185
Sales amount	Total purchase amount last 12 months	8077	13878

Average balance	Average balance in last 12 months	3049	4574
Credit limit	Average credit limit last 12 months	13450	7914
Utilization	Average utilization in last 12 months	0.27	0.32

RESULTS AND FINDINGS

Multivariate regression analyses estimated using ordinary least squares (OLS) is used to analyze the relationship between credit card CLV and its determinants. The SAS GLM procedure was used to run a stepwise regression resulting in models comprised of the strongest predictors of the dependent variable. Even though certain independent variables showed high pair-wise correlation coefficients, variance inflation factors estimated using tolerance statistics from the SAS GLM procedure indicated low levels of multicollinearity. Table 3 can be used to cross-examine the pooled regression with separate regressions for each of the customer segments identified by the bank. Intercepts for models are not displayed to highlight the relative effects of independent variables on predicted customer lifetime values. Further, the Chow test indicated structural breaks by market segment were statistically significant at below the one percent level of significance with the calculated F-statistic equal to 3509 compared to a critical F-statistic of 1.58. As you can see by looking at Table 3, all categories of predictors for all models are statistically significant at below the one percent level. As expected, the effects of independent variables in terms of the sign, size, and level of statistical significance of slope coefficients vary by market segment, even though nearly all variables are statistically significant in the pooled regression. This highlights the importance of taking into consideration how consumer behavior varies by market segment, not to mention the dangers associated with the estimation of a pooled regression that generates results using the entire sample but are not applicable to any customer segment. To further contextualize the results outlined below, the revolver customer segment contributes both interest income and non-interest income to the bank, while contributions from the transactor segment come in the form of non-interest income. Finally, it should be noted that at any time in the future, inactive accounts could become active, at which point their segment will change to either a revolver or transactor based on their specific consumer behavior.

Table 3 Regression Model Estimates

Parameter	M1: all samples	M2: Inactive Accounts	M3: Revolver	M4: Transactor
Average balance	0.08**	0.00	0.07**	0.07**
Utilization rate	114**	0	98**	-88**
Card frequency	0.19**	23.65	0.30**	0.38**
Card sales	0.0044**	0.00	0.0044**	0.0037**
Credit limit	0.0092**	0.0108**	0.0202**	0.0044**
Pre-tax profit	2.06**	-1.05**	1.77**	2.20**
Consumer loan	90**	150**	27*	105**
CD	-216**	-29	-203**	-131**
Checking	-84**	-9	-108**	-53**
Saving	-55**	11	-61**	-37**
Mortgage	-36**	16	-46*	-44**
Marriage:				
Missing	48*	-5	54	-11
Single	-15	-46*	25	-56**
Married	99**	8	175**	23*
Divorced	47**	3	30	1
Separated	N/A	N/A	N/A	N/A
Total bank product				
1 product	-338**	-10	-416**	-203**
2 products	-245**	4	-284**	-148**
3 products	-156**	30	-181**	-85**
4 products	-100**	33	-116**	-41
5 products	-19	59	-34	9
6+ products	N/A	N/A	N/A	N/A
Card recency				
0 month	-6	-104	-158**	45**
1 month	325**	-32	476**	73**
2 months	126**	N/A	217**	-4
3 months	61**	-382	127**	-25
4-6 months	15	-386	53*	-18
7+ months	N/A	N/A	N/A	N/A
Delinquency				
0: current	476**	-2,352**	432**	177**
1-29 DPD	363**	-778	300**	186**
30+ DPD	N/A	N/A	N/A	N/A
Origination FICO				
<600	132**	92**	207**	76**
600-619	75**	151**	172**	-30
620-639	170**	97**	281**	30
640-659	171**	97**	247**	125**
660-679	259**	192**	331**	203**

680-699	248**	173**	316**	180**
700-719	248**	173**	315**	180**
720-739	223**	164**	284**	164**
740-759	191**	143**	266**	124**
760-779	111**	80**	168**	71**
780-799	61**	47**	100**	38**
800+	N/A	N/A	N/A	N/A
Current FICO				
<600	190**	269**	154**	223**
600-619	167**	176**	243**	25
620-639	201**	151**	288**	44
640-659	295**	153**	397**	70**
660-679	356**	112**	467**	153**
680-699	398**	153**	527**	159**
700-719	389**	124**	530**	164**
720-739	358**	155**	484**	160**
740-759	258**	120**	361**	92**
760-779	157**	37**	240**	47**
780-799	45**	-10	118**	-10
800+	N/A	N/A	N/A	N/A
Card Tenure				
<=1YR	565**	319**	582**	560**
>1YR and <=5YR	66**	50**	38*	85**
>5YR and <=10YR	64**	27*	65**	29**
>10YR and <=15YR	17*	12	27	-22**
>15YR and <=20YR	21*	6	27	1
>20YR	N/A	N/A	N/A	N/A
Company Tenure				
<=5YR	-2	3	-55*	-9
>5YR and <=10YR	28**	32*	-19	18
>10YR and <=15YR	52**	10	16	32**
>15YR and <=20YR	18*	-17	-4	1
>20YR and <=30YR	16*	-2	2	-10
>30YR	N/A	N/A	N/A	N/A
Age				
<=20	-205**	-332**	68	-135**
>20 and <=25	57**	-16	228**	97**
>25 and <=30	251**	88**	413**	243**
>30 and <=35	332**	122**	489**	321**
>35 and <=40	415**	130**	602**	342**
>40 and <=45	429**	130**	632**	346**
>45 and <=50	374**	96**	583**	275**
>50 and <=55	291**	86**	489**	195**
>55 and <=65	182**	31*	362**	121**
>65 and <=70	88**	7	236**	55**
>70	N/A	N/A	N/A	N/A
Income Group				
Missing	-110**	94*	-174**	-71**
<=$10k	-87**	173**	-161**	-45*
>$10K and <=$30K	-64**	132*	-127**	-30
>$30K and <=$50K	-8	204**	-65*	-0
>$50K and <=$100K	14	118*	-11	-35*
>$100K	N/A	N/A	N/A	N/A

** P<.001; *P<.05

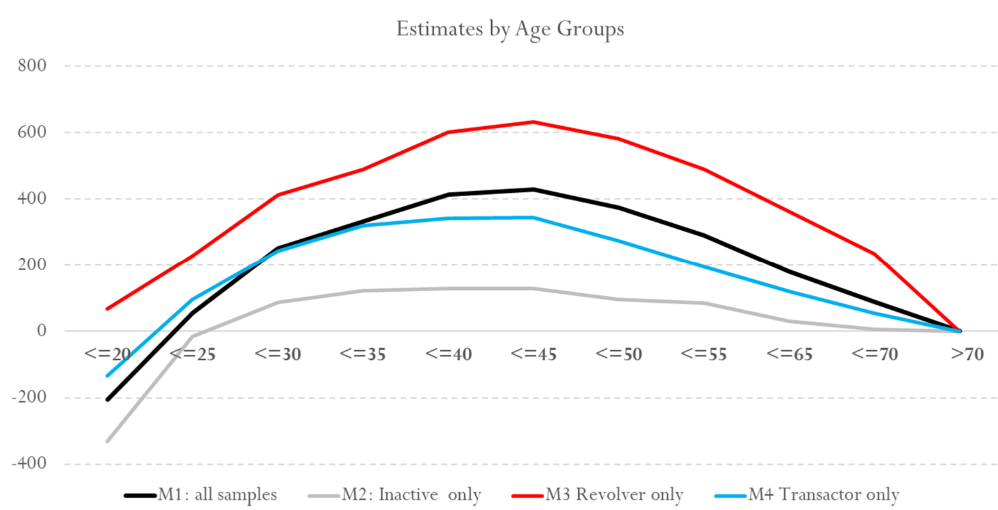
Factors that Impact Credit Card Lifetime Values

Multivariable regression models provide strong evidence that each of the categories of variables are important in modeling customer lifetime value, and that the results vary for from one market segment to the next. Except for gender, demographic variables including age, marital status, income, and credit score are statistically significant at below the one percent level of significance. Company and product relationship variables including company tenure, credit card tenure, the total number of bank products owned, and whether a customer has a checking account, savings account, time deposit, consumer loan, and mortgage are also statistically significant with variable effect sizes and signs that vary by market segment. Credit card account behavioral variables including customer profits, delinquency status, credit card utilization, credit card limit, credit card balance, and the recency, frequency, and monetary value of transactions were all statistically significant at below the one percent level, with varying effect sizes between inactive, revolver, and transactor customer segments.

In the current study, gender was not statistically significant, although support for its inclusion in customer lifetime value models has been provided by Barrios and Lansangan (2012), Costa et al. (2018), Estrella-Ramon et al. (2017), and Haenlein et al. (2007). For each categorical variable included in the analysis, SAS was used to create $n - 1$ dummy variables, where n is equal to the number of distinct groups, for which coefficients were compared to the base group included in the intercept of the model. For categorical variable marriage, married customers had the largest coefficient, equal to 99 in the pooled sample, compared to a value of -15 for single customers, 47 for divorced customers, and 0 for customers who were separated from their spouse. This result contradicts a study by Chang and Ijose (2016) that examined the effect of whether a customer was married on the logged value of credit card purchases.

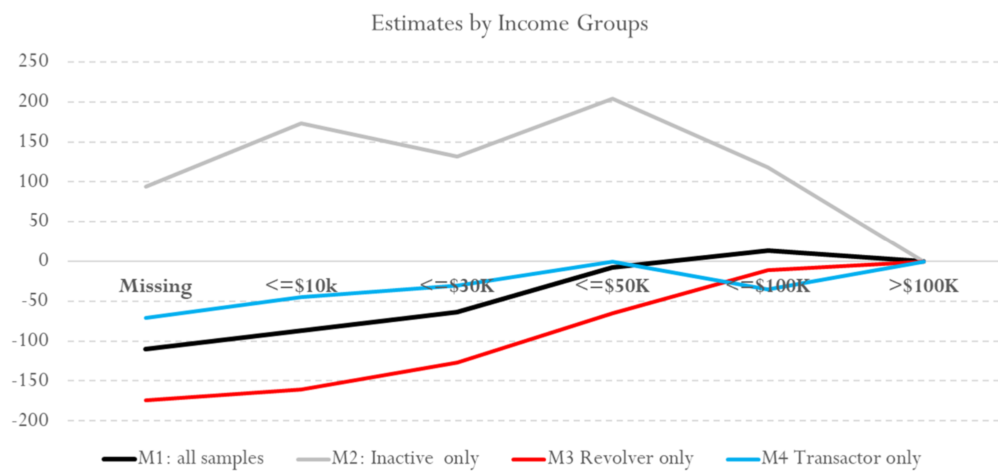
As can be seen by looking at figure 1 below, customer age has a reversed u-shaped relationship with CLV, which is consistent with expectations. Younger customers (<=30) have less purchasing power and older customers (>55) have a lower level of demand for goods and services. Customers with age greater than 35 and less than or equal to 45 have the highest CLV (with a coefficient of 415 for the pooled sample). Chang and Ijose (2016) used customer age, customer age squared, and a customer age and income interaction term in a model that examined the effect of age on the amount of credit card purchases. A negative relationship between credit card purchases and age was found; however, it should be noted that credit card purchases are fundamentally different than customer lifetime values.

Figure 1 The Effect of Age on CLV



Generally, higher incomes are correlated with higher CLV values. As you can see by looking at figure 2 below, customer income has an increasing monotonic relationship with CLV – i.e., higher customer income levels are associated with higher customer lifetime values. For the "Inactive" market segment, those who earn between $50K and $100K have CLV values that are 118 units higher than those earning more than $100K. Other income categories can be interpreted similarly. Customers earning $30K to $50K have CLV values that are 204 units higher than those earning more than $100K; hence the effect is positive relative to the base category. Relative to those earning more than $100K, CLV values for all other income categories are higher, at least for inactive customers; hence, CLV is lowest for those earning > $100K. For other market segments, CLV values are generally lower relative to the base category of $100K, and the effects of being in different income categories vary relative to the base category.

Figure 2 The Effect of Income on CLV



Figures 3 and 4 provide a graphical representation of the effect that customer credit scores have on CLV. Results are provided for both the pooled regression and each customer segment. the magnitude of the effect of customer credit scores varies by market segment with revolvers having the highest effect sizes. As with customer age, both origination customer credit scores and current customer credit scores can be described as having a reversed U-shaped relationship with CLV. Further, the effect of OFICO is small for low- and high-credit customers, but higher for customers with credit scores in between 660 and 759. Near prime customers with origination FICO scores between 660 and 679 and current FICO between 680 and 699 are associated with higher CLV values. These customers were more likely to be revolvers with current balances. High FICO customer groups tended to be comprised of either inactive customers, possibly due to their having more credit card options relative to customer groups with lower credit scores, or the less profitable transactor customer segment.

Figure 3 The Effect of Original FICO on CLV

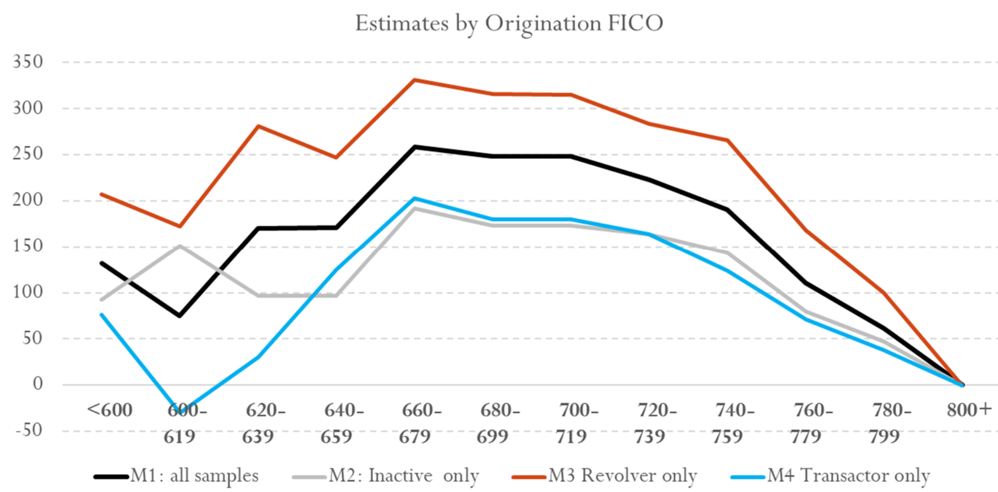
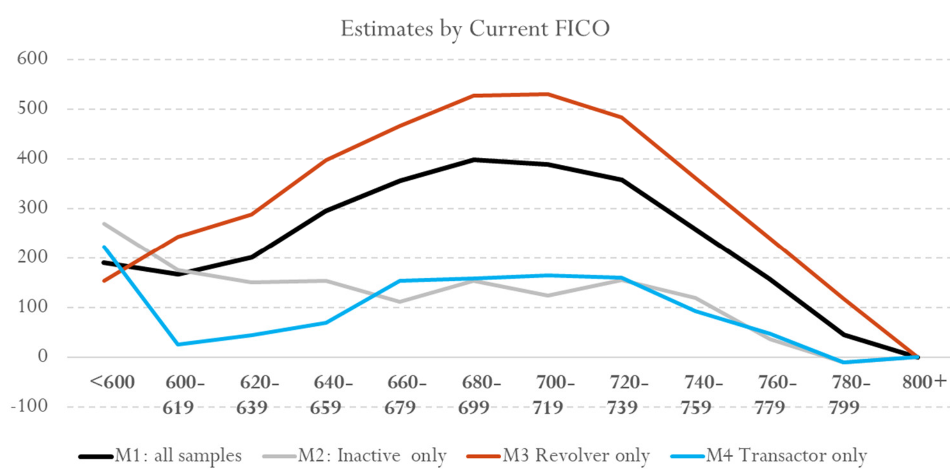
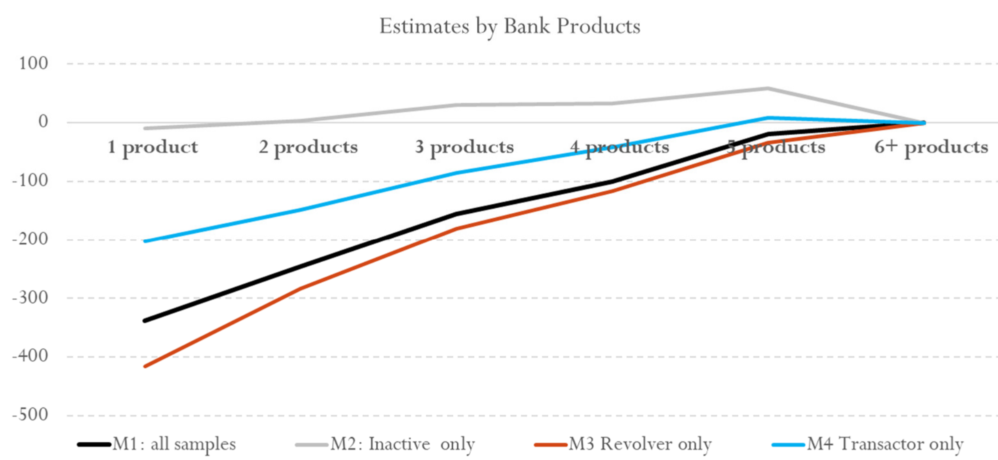


Figure 4 The Effect of Current FICO on CLV



The regression coefficients for all customer product relationship variables were statistically significant at below the one percent level of significance. Overall, the total number of bank products customers own is positively related to CLV. Although coefficients for product categories are negative, the effect sizes become less negative as the number of bank products customers own increases. Of course, this is determined by the reference category chosen for the categorical variable. It was hypothesized that card tenure matters, and it does. However, researchers expected a positive relationship with CLV. The results showed the opposite. Recent card holders have higher CLV values relative to tenured customers. More specifically, company tenure showed a reversed U-shape relationship with CLV. Customers with tenure between 10 and 15 years had the highest long-term profitability. Another unexpected finding deals with the relationship between CLV and the ownership of certain bank products. For example, customers with depository accounts and mortgages had lower customer lifetime values.

Figure 5 The Effect of Bank Product Ownership and CLV



As can be seen by looking at the regression output above, credit card account behavior variables were all statistically significant at below the one percent level of significance. In the pooled model, Average balance, Credit limit, Utilization rate, Card frequency, Card sales, and Pre-tax profit are positively related to CLV, which is as expected. Recency is positively related to CLV. For the pooled sample, whether a customer had a transaction 1 month ago is associated w/ a 325 unit increase in CLV. As the number of months since the last transaction increases, the effect size on CLV decreases for the pooled sample, Revolvers, and Transactors. Also as expected, current customers are more profitable than those who are delinquent. The Utilization rate for "Transactors" is negatively related to CLV. One possible reason is that customers in the "Transactor" segment pay-off their balances each month, and therefore incur very low borrowing fees. This causes the bank to incur negative interest income. Also, the interchange income from the "Transactor" category is low compared to interest income. Further, 95% of members in the "Inactive" segment have negative profits.

CONTRIBUTIONS AND LIMITATIONS

The purpose of this study was to address a series of gaps in literature that examines credit card customer lifetime values. Long-term profitability data was used to construct a series of multivariate statistical models that explore the determinants of credit card customer lifetime value. This is the first paper to include more than 5 years of data, which made it possible to obtain more accurate predictions. A unique sample of 500,000 credit card customers randomly sampled from a large US bank made it possible to assess the relationship between CLV and three categories of explanatory variables, many of which are often excluded from credit card CLV models, including customer demographic characteristics, company relationship variables, and variables that measure credit card account behavior. A general linear regression model comprised of the full sample was used to cross-examine the effects of the determinants of CLV by market segment. Statistically significant findings were obtained for all models estimated, with customer segmentation models outperforming the results from the pooled regression.

Since the models were developed using data from only one US bank, researchers are encouraged to replicate these results in CLV analyses of other banks, both domestic and abroad. Furthermore, the models estimated in the current research used cross-section versus longitudinal data. Researchers are advised to construct panel data to assess the degree to which the effects of the determinants of CLV vary over time through the inclusion of time-varying covariates – e.g., macro-economic variables or firm-decision variables. Although this study is one of the first to construct customer a model of credit card CLV using a wide variety of indicators, it did not include customer perception variables – e.g., consider satisfaction, or a net promotion score. It is also advised that researchers construct the dependent variable using profitability data obtained from a portfolio of products. Finally, the current research modeled CLV directly. It might be worthwhile to first model CLV components, for example the customer retention rate, before using those results to estimate CLV.

REFERENCES

Aeron, H., Bhaskar, T., Sundararajan, R., Kumar, A., & Moorthy, J. (2008). A metric for customer lifetime value of credit card customers. *Journal of Database Marketing & Customer Strategy Management*, 15(3), 153-168.

Alizadeh Zoeram, A., & Karimi Mazidi, A. R. (2018). New approach for customer clustering by integrating the LRFM model and fuzzy inference system. *Iranian Journal of Management Studies*, 11(2), 351-378.

Barrios, E., & Lansangan, J. R. (2012). Forecasting Customer Lifetime Value: A Statistical Approach. *Philippine Management Review*, 19, 23-34.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of interactive marketing*, *12*(1), 17-30.

Calciu, M. (2009). Deterministic and stochastic Customer Lifetime Value models. Evaluating the impact of ignored heterogeneity in non-contractual contexts. *Journal of Targeting, Measurement and Analysis for Marketing,* 17(4), 257-271.

Canals-Cerda J & Kerr S. (2015), Forecasting credit card portfolio losses in the great recession: a study in model risk, *Journal of Credit Risk*, 2015, FRB Of Philadelphia Working Paper No. 14-10

Casas-Arce, P., Martínez-Jerez, F. A., & Narayanan, V. G. (2017). The impact of forward-looking metrics on employee decision-making: The case of customer lifetime value. *The Accounting Review*, 92(3), 31-56.

Chang, C. W., & Ijose, O. (2016). Measuring customer lifetime value: An application in credit card industry. *Academy of Business Research Journal*, 1, 7.

Costa, S. B. D., Claro, D. P., & Bortoluzzo, A. B. (2018). The influence of customer value on word of mouth: a study of the financial services company in Brazil. *Review of Business Management*, 20, 210-231.

Däs, M., Klier, J., Klier, M., Lindner, G., & Thiel, L. (2017). Customer lifetime network value: customer valuation in the context of network effects. *Electronic markets*, 27(4), 307-328.

Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007). Modeling CLV: A test of competing models in the insurance industry. *Quantitative Marketing and Economics,* 5(2), 163-190.

Dwyer, F. R. (1989). Customer lifetime valuation to support marketing decision making. *Journal of direct marketing*, 3(4), 8-15.

Ekinci, Y., Uray, N., & Ülengin, F. (2014). A customer lifetime value model for the banking industry: a guide to marketing actions. *European Journal of Marketing*, 48 (3/4), 761-784.

Estrella-Ramón, A., Sánchez-Pérez, M., Swinnen, G., & VanHoof, K. (2017). A model to improve management of banking customers. *Industrial Management & Data Systems*, 117(2), 250-266.

Finlay S. (2008), Towards profitability: a utility approach to the credit scoring problem, *Journal of Operational Research Society*, 59, 921-931, doi:10.1057/palgrave.jors.2602394

Gao, L., Melero-Polo, I., & Sese, F. J. (2020). Customer equity drivers, customer experience quality, and customer profitability in banking services: The moderating role of social influence. *Journal of Service Research*, 23(2), 174-193.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker N., & Sriram, S. (2006). Modeling customer lifetime value. *Journal of Service Research*, 9(2), 139-155.

Haenlein, M., Kaplan, A. M., & Beeser, A. J. (2007). A model to determine customer lifetime value in a retail banking context. *European Management Journal*, 25(3), 221-234.

Han, S. H., Lu, S. X., & Leung, S. C. (2012). Segmentation of telecom customers based on customer value by decision tree model. *Expert Systems with Applications*, 39(4), 3964-3973.

Ho, T. H., Park, Y. H., & Zhou, Y. P. (2006). Incorporating satisfaction into customer value analysis: Optimal investment in lifetime value. *Marketing Science*, 25(3), 260-277.

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2019). Comparative analysis of selected probabilistic customer lifetime value models in online shopping. *Journal of Business Economics and Management*, 20(3), 398-423.

Kou G., Peng Y., Shi Y., Wise M., and Xu W.X. (2005), Discovering credit cardholders' behavior by multiple criteria linear programming, *Annuals of Operations Research,* 135, 261-274, 2005

Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing,* 80(6), 36-68.

Li, C. Q., Marcel, W., & Gao, T. (2012). Does A Long-Term and Active Customer of A Current Savings Account Have A High Lifetime Value? *Contemporary Management Research*, 8(2), 141-160

Malthouse, E. C., & Blattberg, R. C. (2005). Can we predict customer lifetime value? *Journal of interactive marketing*, 19(1), 2-16.

Memarpour, M., Hassannayebi, E., Fattahi Miab, N., & Farjad, A. (2019). Dynamic allocation of promotional budgets based on maximizing customer equity. *Operational Research*, 21(4), 2365-2389.

Oblander, E. S., Gupta, S., Mela, C. F., Winer, R. S., & Lehmann, D. R. (2020). The past, present, and future of customer management. *Marketing Letters*, 31(2), 125-136.

Osipenko, D. (2018). Investigation into methods of predicting income from credit card holders using panel data.

Pulina, M. (2011). Consumer behaviour in the credit card market: a banking case study. *International Journal of Consumer Studies*, 35(1), 86-94.

Rezaei, M., Sanayei, A., Amiri Aghdaie, S. F., & Ansari, A. (2022). Improving the Omnichannel customers' lifetime value using association rules data mining: A case study of Agriculture Bank of Iran. *Iranian Journal of Management Studies*, 15(1), 49-68.

So M.C., Thomas L.C., Seow H.V., Mues C. (2014), Using a transactor/revolver scorecard to make credit and pricing decisions, *Decision Support Systems*, 59, 143-151. DOI: 10.1016/j.dss.2013.11.002

Stewart R.T. (2011), A profit-based scoring system in consumer credit: making acquisition decisions for credit cards, *Journal of the Operational research Society*, 62, 1719-1725

Till R.J. & Hand D.J. (2003), Behavior models of credit card usage, *Journal of Applied Statistics,* 30 (10), 1201-1220

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of marketing,* 68(4), 106-125.

Yoo, M., Bai, B., & Singh, A. (2020). The evolution of behavioral loyalty and customer lifetime value over time: investigation from a Casino Loyalty Program. *Journal of Marketing Analytics,* 8(2), 45-56.