



A modified Pareto/NBD approach for predicting customer lifetime value

Nicolas Glady, Bart Baesens and Christophe Croux

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

A Modified Pareto/NBD Approach for Predicting Customer Lifetime Value

Nicolas Glady^a Bart Baesens^{a,b} Christophe Croux^{a *}

^a Faculty of Economics and Management, K.U.Leuven

^b School of Management, University of Southampton

Abstract

Valuing customers is a central issue for any commercial activity. The customer lifetime value (CLV) is the discounted value of the future profits that this customer yields to the company. In order to compute the CLV, one needs to predict the future number of transactions a customer will make and the profit of these transactions. With the Pareto/NBD model, the future number of transactions of a customer can be predicted, and the CLV is then computed as a discounted product between this number and the expected profit per transaction. Usually, the number of transactions and the future profits per transaction are estimated separately. This study proposes an alternative. We show that the dependence between the number of transactions and their profitability can be used to increase the accuracy of the prediction of the CLV. This is illustrated with a new empirical case from the retail banking sector.

*We thank ING Belgium for their support and useful information, especially Martine George head of the customer intelligence department. We also thank Bruce Hardie for kindly providing the CDNOW data set. All correspondence should be sent to the first author Nicolas Glady, Naamsestraat 69, B-3000 Leuven; +32.16.326817; Nicolas.Glady@econ.kuleuven.ac.be.

1 Introduction

Valuing customers is a central issue of any commercial activity. The value of an individual customer is important for the detection of the most valuable ones, which deserve to be closely followed, and for the detection of the less valuable ones, to which the company should pay less attention. At the aggregated level, a marketing campaign targeting a group of customers can be budgeted more efficiently when the value of this group is known. Customers are an important asset, and as such, have to be precisely valued.

Customer valuation has been discussed by several papers in the customer relationship management literature, for example Dwyer [1997], Berger and Nasr [1998], Rust et al. [2004] and Malthouse and Blattberg [2005]. The value of a customer has long been defined with regard to the longevity of his/her historical financial value. However, Reinartz and Kumar [2000] criticized this method, since they demonstrated that a long life-cycle and the profitability of a customer were not necessarily related. On the opposite, Rust et al. [2004] emphasized that marketing strategy should focus on projected future financial return using the total value of the customer base. Supporting this idea, Gupta et al. [2004] showed that the earnings of a company, and hence its value, are a function of the total *Customer Lifetime Value* (CLV). The CLV is defined as the discounted value of the future profits yielded by a customer to the company. The issue is to predict the future profits when the timing and the profit of future transactions is not known, that is in a non-contractual setting, as discussed in Mulhern [1999] and Bell et al. [2002].

The Pareto/NBD model, introduced by Schmittlein et al. [1987], is referred by several authors [e.g. Mulhern, 1999, Niraj et al., 2001, Jain and Singh, 2002] as a powerful technique to predict the future activity of a customer in a non-contractual relationship. For examples of implementation, see Schmittlein and Peterson [1994], Reinartz and Kumar [2000] and Fader et al. [2005a]. Since the Pareto/NBD model forecasts only the probability of activity and the number of transactions of a customer, some adaptations are to be made in order to incorporate the profit of the transactions and to estimate the CLV. Herefore, the Gamma/Gamma submodel is used, as in Fader et al. [2005a]. A key assumption, made by this Pareto/NBD-based model for CLV prediction, is the independency between the number of transactions of a customer and the related profit per transaction. We will propose a modified model, not relying on this independence assumption.

The purpose of our paper is to propose a modified Pareto/NBD-based approach for the CLV prediction. In our empirical study, using customer stock exchange orders provided by a retail banker, we will show that the newly proposed method has better forecasting performance than the tradi-

tional Pareto/NBD model, and that it also outperforms a standard regression approach.

Our paper is organized as follows: after an introduction of the concept of customer lifetime value in Section 2, we present in Section 3 the Pareto/NBD submodel previously used to predict the CLV. Section 4 proposes a modification, which we call the Pareto/Dependent model. Section 5 presents then the empirical application, where the predictive performance of the different models is compared.

2 Customer Lifetime Value Definition and Principles

Nowadays one can see a proliferation of valuation methods using both terms of “Customer Lifetime Value” or “Customer Equity”. For an overview, see Pfeifer et al. [2005]. This paper follows Gupta et al. [2004], defining the value of a customer as the expected sum of discounted future earnings, where a customer generates a profit margin for each period.

The CLV is a function of all the transactions an individual customer will make in the future. Strictly speaking, all future transactions should be considered. Nevertheless, in order to compare our predictions with actual data, we will work with a finite horizon h . Consequently, the CLV of the customer i for the horizon h is

$$\text{CLV}_{i,h} = \sum_{k=1}^h \frac{\text{CF}_{i,k}}{(1+d)^k}, \quad (2.1)$$

where d is the discount rate, assumed to be constant. Hence it is the sum of discounted net cash flows, where $\text{CF}_{i,k}$ is the net cash flow (i.e. the total gains less the total costs) due to the activity of customer i during the time period k .¹ The CLV of a customer is obviously changing over time. Nevertheless, we will not introduce this time dependency in the notation, since in our empirical study the moment of prediction of the CLV is identical for all customers.

This paper studies the prediction of $\text{CLV}_{i,h}$ given the past purchases information. The optimal prediction, in the least square sense, is

$$\widehat{\text{CLV}}_{i,h} = E[\text{CLV}_{i,h} | \text{Past Purchase Information}]. \quad (2.2)$$

For a recent review on CLV modeling, see Gupta et al. [2006]. Most papers predicting the CLV [e.g. Schmittlein and Peterson, 1994, Venkatesan and

¹For simplicity purposes, we consider time periods of equal length and the discount is computed as if all cash flows were obtained end-of-period.

Kumar, 2004, Fader et al., 2005a] are using a two step scheme. First they forecast the future number of transactions of each individual. For instance Fader et al. [2005a], studying the sales of a CD retailer, first predict the future number of CD purchases. Then, the individual average profit per transaction is estimated. These values are estimated at the customer level and, if the product of the future number of transactions and the profit per transaction is discounted and summed up, it yields an approximation of the CLV for each customer. Section 3 reviews this approach. Section 4 modifies this approach, motivated by the empirical evidence that the number of transactions and the average profit per transaction are not independent of each other.

3 Customer Lifetime Value Prediction

This Section will describe the common approach for predicting the CLV. First the number of transactions in the future is predicted using the Pareto/NBD submodel (described in the next subsection). Next, another submodel provides an estimate of the average profit per transaction. The CLV is then computed as a discounted product of the future number of transactions and the average profit per transaction. Such an approach is what we call a Pareto/NBD-based model.

3.1 Pareto/NBD Submodel Description

We will describe the Pareto/NBD submodel as proposed by Schmittlein et al. [1987]. Based on the past observations, the parameters of the Pareto/NBD submodel are estimated. Then, one is able to forecast the future activity of a customer. All predictions are made at the same point in time, i.e. the present time or “now”.

Three past purchasing behavior measures are required for every customer. The first purchasing information is the *cohort* T_i . This is the time between the entry of the individual i as a customer of the company until now. If we denote $x_{i,k}$, the number of transactions the customer i has made after k time units, then this customer has made x_{i,T_i} transactions until the present time. The latter value is called the *frequency* and is the second purchasing variable required by the Pareto/NBD submodel. We use the shorthand notation $x_i \equiv x_{i,T_i}$ throughout the paper. Note that the total number of transactions from the beginning of the relationship between the customer and the company until “now” is $x_i + 1$, because the first transaction occurs at the moment of entry. The third and last purchasing information required by the Pareto/NBD submodel is the time between the entry date and the

last purchase date. This value is called the *recency* and is denoted t_i . The more recent is the last purchase, the higher t_i will be, and $0 < t_i \leq T_i$. The purchasing information on customer i contains thus three observed values: T_i , x_i and t_i respectively the cohort, the frequency and the recency, for each $1 \leq i \leq n$, where n is the total sample size.

In the context of the Pareto/NBD submodel, a customer is said to be active as long as this customer is making transactions. Once inactive, a customer will not make any transactions anymore. There are five assumptions to be made about the purchase event process and the time that a customer stays active. First, while active, a customer i makes purchases according to a Poisson process with rate λ_i . We will denote this assumption as (A1). Let the (unobserved) time at which the customer i becomes inactive be denoted by τ_i . If the customer i is still active at T_i (so $\tau_i > T_i$), the number of purchases x_i in $(0, T_i[$ has the Poisson distribution

$$P[x_i | \lambda_i, \tau_i > T_i] = e^{-\lambda_i T_i} \frac{(\lambda_i T_i)^{x_i}}{x_i!}. \quad (3.1)$$

The second hypothesis (A2) is that each customer remains active during a time being exponentially distributed with death rate μ_i

$$f(\tau_i | \mu_i) = \mu_i e^{-\mu_i \tau_i}. \quad (3.2)$$

Since the parameters λ_i and μ_i can be different among customers, the Pareto/NBD submodel makes three assumptions on the heterogeneity across customers. The assumption (A3) is that the purchasing rate λ_i for the different customers is distributed according to a Gamma distribution across the population of customers.

$$g(\lambda_i | r, \alpha) = \frac{\alpha^r}{\Gamma(r)} \lambda_i^{r-1} e^{-\alpha \lambda_i}; \quad r, \alpha > 0, \quad (3.3)$$

with $E[\lambda_i | r, \alpha] = r/\alpha$. The fourth assumption (A4) is that the death rates μ_i are distributed according to a different Gamma distribution across customers,

$$g(\mu_i | s, \beta) = \frac{\beta^s}{\Gamma(s)} \mu_i^{s-1} e^{-\beta \mu_i}; \quad s, \beta > 0, \quad (3.4)$$

with $E[\mu_i | s, \beta] = s/\beta$. Finally, the purchasing rates λ_i and the death rate μ_i are considered as distributed independently of each other (A5). The population parameters r , α , s and β are unknown and need to be estimated.

In this paper, the population parameters r , α , s and β are estimated by Maximum Likelihood (MLE). In Fader and Hardie [2005] the likelihood for

an individual i with purchase history (x_i, t_i, T_i) , $L_i = L(r, \alpha, s, \beta | x_i, t_i, T_i)$, is shown to be equal to

$$L_i = \frac{\Gamma(r + x_i) \alpha^r \beta^s}{\Gamma(r)} \left\{ \frac{1}{(\alpha + T_i)^{r+x_i} (\beta + T_i)^s} + \left(\frac{s}{r + s + x_i} \right) A_0 \right\}, \quad (3.5)$$

with, for $\alpha \geq \beta$,²

$$A_0 = \frac{F(a_i, b; c_i; z(t_i))}{(\alpha + t_i)^{r+s+x_i}} - \frac{F(a_i, b; c_i; z(T_i))}{(\alpha + T_i)^{r+s+x_i}}, \quad (3.6)$$

where,

$$a_i = r + x_i + s; \quad b = s + 1; \quad c_i = r + x_i + s + 1; \quad z(y) = \frac{\alpha - \beta}{\alpha + y}, \quad (3.7)$$

and F is the Gaussian hypergeometric function³, which is a power series of the form

$$F(a, b; c; z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!},$$

where $(a)_j$ is the Pochhammer's symbol, which denotes the ascending factorial, $a \times (a + 1) \dots \times (a + j - 1)$. An alternative to the MLE is to use the method-of-moments, but Reinartz and Kumar [2003] showed that the method-of-moments yields similar results.

Once the parameters are estimated, one can estimate the probability for a customer to be alive, and predict the future number of transactions this customer will make. The estimated conditional probability, $P[\tau_i > T_i | x_i, t_i, T_i]$, for the customer i of being active at the present moment, given the frequency, recency and cohort of this customer is

$$\hat{p}_i = \frac{1}{1 + \frac{\hat{s}}{\hat{r} + x_i + \hat{s}} \left[\left(\frac{\hat{\alpha} + T_i}{\hat{\alpha} + t_i} \right)^{\hat{r} + x_i} \left(\frac{\hat{\beta} + T_i}{\hat{\alpha} + t_i} \right)^{\hat{s}} F(\hat{a}, \hat{b}; \hat{c}; \hat{z}(t_i)) - \left(\frac{\hat{\beta} + T_i}{\hat{\alpha} + T_i} \right)^{\hat{s}} F(\hat{a}, \hat{b}; \hat{c}; \hat{z}(T_i)) \right]}. \quad (3.8)$$

The submodel also provides the (unconditional) expected value of the number of transactions a customer will make over time,

$$E[x_{i,k}] = \frac{\hat{r} \hat{\beta}}{\hat{\alpha} (\hat{s} - 1)} \left[1 - \left(\frac{\hat{\beta}}{\hat{\beta} + k} \right)^{\hat{s} - 1} \right]. \quad (3.9)$$

²For $\alpha < \beta$ the function is slightly different. See Fader and Hardie [2005] for more details.

³The standard reference for the Gaussian hypergeometric function is Abramowitz and Stegun [1972].

However, for the prediction of the CLV of an individual customer i , one would rather need to estimate $E[x_{i,T_i+k}|x_i, T_i, t_i]$. That is the conditional expectation of the number of transactions a customer i makes until the end of time period T_i+k , given the recency, frequency and cohort of this customer. As shown in Fader and Hardie [2005], an estimate of this quantity is

$$\hat{x}_{i,T_i+k} = x_i + \frac{\Gamma(\hat{r} + x_i) \hat{\alpha}^{\hat{r}} \hat{\beta}^{\hat{s}}}{\Gamma(\hat{r})(\hat{\alpha} + T_i)^{\hat{r}+x_i} (\hat{\beta} + T_i)^{\hat{s}} \hat{L}_i} \frac{(\hat{r} + x_i)(\hat{\beta} + T_i)}{(\hat{\alpha} + T_i)(\hat{s} - 1)} \left[1 - \left(\frac{\hat{\beta} + T_i}{\hat{\beta} + T_i + k}\right)^{\hat{s}-1}\right], \quad (3.10)$$

where $\hat{L}_i = L(\hat{r}, \hat{\alpha}, \hat{s}, \hat{\beta}|x_i, t_i, T_i)$ is the estimated likelihood of equation (3.5) and $\Gamma(\cdot)$ denotes the standard Gamma function. The expected number of transactions during the k^{th} future time period is then $\hat{x}_{i,T_i+k} - \hat{x}_{i,T_i+k-1}$.

An easier method for implementation has been proposed by Fader et al. [2005b], the Beta-geometric/NBD submodel, assuming a beta-geometric distribution instead of a Pareto distribution. Changing only slightly the assumptions of the Pareto/NBD submodel, this method can even be implemented in Excel. Nevertheless, in this paper, we will use the traditional Pareto/NBD owing to its usage in the reference literature.

3.2 The Gamma/Gamma Submodel

Until now, the profit of a transaction, needed to predict the CLV, was not introduced in the model yet. For this purpose, the Gamma/Gamma submodel of Fader et al. [2005a] can be taken. This submodel estimates the average profit per transaction of a customer. The profit of a transaction is then defined as the net cash flow it yields. We denote $z_{i,1}, \dots, z_{i,x_i}$, the profit of each observed transaction made by the customer i , and $m_{i,k}$ the average profit of the transactions of the customer i from the beginning of the customer relationship with the company until time k . Note that when $k = T_i$,

$$\tilde{m}_i \equiv m_{i,T_i} = \sum_{l=1}^{x_i} z_{i,l} / x_i.$$

This is the monetary value, the average profit per transaction of a customer until now.

An assumption made in Fader et al. [2005a] (denoted A6) is that, for each individual i , the profit per transaction is independent of the number of transactions. Moreover, it is assumed in the Gamma/Gamma submodel that the expected profit per transaction does not vary over time and we denote it by m_i . The monetary value \tilde{m}_i is then a sample estimate of m_i . A further assumption (A7) is that the $z_{i,l}$ are Gamma distributed with shape

parameter px_i and scale parameter $1/\nu_i$. The last assumption (A8) is that the values of ν_i are again Gamma distributed across the population with shape parameter q and scale parameter $1/\gamma$. It leads to the total likelihood of the Gamma/Gamma submodel, depending on three unknown parameters

$$L(p, q, \gamma) = \prod_{i=1}^n \left(\frac{\Gamma(px_i + q)}{\Gamma(px_i)\Gamma(q)} \frac{\gamma^q \tilde{m}_i^{px_i-1} x_i^{px_i}}{(\gamma + \tilde{m}_i x_i)^{px_i+q}} \right). \quad (3.11)$$

Finally, once these parameters have been estimated by MLE, the conditional expectation of the average profit per transaction for a customer i is estimated as

$$\hat{m}_i = \left(\frac{\hat{q} - 1}{\hat{p}x_i + \hat{q} - 1} \right) \frac{\hat{\gamma}\hat{p}}{\hat{q} - 1} + \left(\frac{\hat{p}x_i}{\hat{p}x_i + \hat{q} - 1} \right) \tilde{m}_i. \quad (3.12)$$

This is a weighted average of the estimated population mean of the profit $(\hat{\gamma}\hat{p})/(\hat{q} - 1)$, and \tilde{m}_i . For more details on the Gamma/Gamma submodel, we refer to Fader et al. [2005a].

3.3 Description of the Pareto/Independent Model

Most papers, such as Schmittlein and Peterson [1994], Venkatesan and Kumar [2004] or Fader et al. [2005a], apply the same principle to predict the CLV. Once the number of transactions in a future time period k is estimated using (3.10), they multiply this value by the expected average profit per transaction given by (3.12). It follows that the CLV is estimated as

$$\widehat{\text{CLV}}_{i,h} = \sum_{k=1}^h \frac{(\hat{x}_{i,T_i+k} - \hat{x}_{i,T_i+k-1})\hat{m}_i}{(1+d)^k}. \quad (3.13)$$

This prediction requires the recency (t_i via equation 3.10), the frequency (x_i via equation 3.10) and the monetary value (m_{i,T_i} via equation 3.12). It fits within the well-known RFM (recency, frequency and monetary) framework.

During our empirical study, when predicting the value of the CLV, we will first apply the method described in this section. This model is referred to as the Pareto/Independent model. In the following, we mitigate the assumption that the average profit per transaction can be estimated independently of the number of transactions. The resulting modified approach will be referred to as the Pareto/Dependent model and is outlined in the next section.

4 A Modified Pareto/NBD Approach for CLV Prediction

In the empirical application we will discuss in Section 5, we will show that the independence assumption (A6) between the number of transactions and the average profit per transaction is questionable. In Fader et al. [2005a], the authors found an average value of .06 for the correlation between \tilde{m}_i and x_i in their empirical application. They accept the independence hypothesis nevertheless, arguing that this value is very small. The new approach takes into account a possible dependency between the number of transactions and the average profit per transaction. This dependency will be designed at the customer level, accounting for the heterogeneity in the population. Moreover, we do not require a constant expected profit per transaction over time. The resulting model will be referred to as the Pareto/Dependent model.

Let us assume that the number of transactions and the average profit per transaction of a customer i are related by the model

$$\log\left(\frac{m_{i,k}}{E[m_{i,k}]}\right) = r_i \log\left(\frac{x_{i,k}}{E[x_{i,k}]}\right) + \varepsilon_i. \quad (4.1)$$

Here r_i is a coefficient of dependence for which an estimation method will be provided in the next paragraph. Equation (4.1) links the deviation of the observed from the expected average profit per transaction, with the same deviation for the number of transactions. Here, the expected values are provided by the Pareto/NBD submodel for the number of transactions and by the Gamma/Gamma submodel for the monetary values, using equation (3.9) and the quantity $(\hat{\gamma}\hat{p})/(\hat{q} - 1)$ respectively.

The idea is that the monetary value of a customer depends on the number of transactions he/she is making. This dependency can be different across customers. Customers having a high number of future transactions may have a high monetary value, but the opposite could also happen. Therefore we model the dependency coefficient r_i as a function of explicative variables, for which we take the cohort, recency and the probability of being an active customer. The latter variable is relevant because the more likely a customer is expected to remain active, the more likely this customer will have positive dependency between the average profit per transaction and its number of transactions. Hence,

$$r_i = \alpha_1 \hat{p}_i + \alpha_2 T_i + \alpha_3 t_i + \alpha_0. \quad (4.2)$$

Taking (4.1) for $k = T_i$ and inserting (4.2) yields

$$\begin{aligned} \log\left(\frac{\tilde{m}_i}{E[m_{i,k}]}\right) &= \alpha_1 \hat{p}_i \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \alpha_2 T_i \log\left(\frac{x_i}{E[x_{i,k}]}\right) \\ &+ \alpha_3 t_i \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \alpha_0 \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \varepsilon_i, \end{aligned} \quad (4.3)$$

for $1 \leq i \leq n$. Estimating this regression equation yields estimates for the parameter α_0 , α_1 , α_2 and α_3 , and hence also an estimate for the r_i .

The average profit per transaction in the period $[0, T_i + k]$ can then be estimated as well. From equation (4.1) it follows that

$$\log\left(\frac{m_{i,T_i+k}}{E[m_{i,k}]}\right) - \log\left(\frac{\tilde{m}_i}{E[m_{i,k}]}\right) = r_i \log\left(\frac{x_{i,T_i+k}}{E[x_{i,T_i+k}]}\right) - r_i \log\left(\frac{x_i}{E[x_{i,T_i}]}\right).$$

This yields as prediction of m_{i,T_i+k} ,

$$\hat{m}_{i,T_i+k} = \tilde{m}_i \left(\frac{\hat{x}_{i,T_i+k}/E[x_{i,T_i+k}]}{x_i/E[x_{i,T_i}]} \right)^{\hat{r}_i}, \quad (4.4)$$

with x_i the observed number of transactions in the past (frequency), T_i the cohort of customer i and \hat{x}_{i,T_i+k} , given by (3.10), as prediction for x_{i,T_i+k} . One can see that \hat{m}_{i,T_i+k} is now a function of the time period k . Then, by (2.1), the CLV of customer i , computed for an horizon of h periods, is estimated under the Pareto/Dependent model by

$$\widehat{\text{CLV}}_{i,h} = \sum_{k=1}^h \frac{\hat{x}_{i,T_i+k} \hat{m}_{i,T_i+k} - \hat{x}_{i,T_i+k-1} \hat{m}_{i,T_i+k-1}}{(1+d)^k}. \quad (4.5)$$

5 Empirical Application

Our first empirical application uses a new data set, provided by a Belgium retail banker, and described in Section 5.1. In Section 5.2 we empirically show that there is a significant correlation between the number of transactions and the profit per transaction for this data set. As a second empirical application, we study the prediction for the CLV of the CDNOW data set, which has been used in Fader et al. [2005a] and serves as a benchmark data set. The accuracy of the prediction of the CLV on these data sets, using different models, is compared. In Section 5.4, it will turn out that for the retail banker data set, the newly proposed Pareto/Dependent model performs best. For the CDNOW data set, where the correlation between the number of transactions and the average profit per transaction is less important, all methods considered perform comparably

5.1 Description of the Retail Banker Data Set

The retail banker data set is provided by a Belgian financial service institution. The data set contains the customers stock exchange transactions data from January 2000 till December 2005. The customers considered were inhabitants of Belgium with no professional activity in the brokerage business. The total number of transactions was 11068877, made by 460566 customers.

These customers could have purchased (or sold) stocks, bonds, mutual funds, derivatives etc.⁴ The profit of a transaction is computed, by a business rule, as a margin of 1% of the amount exchanged at the transaction. When computing the CLV, we will work with monthly time periods. The discount rate is taken as the weighted average cost of capital disclosed in the 2004 financial statement of the Belgian financial service institution, 8.92% yearly, giving a monthly discount rate of $d = 0.7146\%$.

On the total base of customers, we select for the CLV prediction those who made their first transaction between January 2001 and December 2003. The CLV prediction will be made at January 1st, 2004. The remaining two years of data are kept out-of-sample for the model assessment. We consider eight groups of customers for the purpose of comparison. Each group (later called cohort) is composed of customers who started their relationship (date of first purchase) during the same quarter. For instance, the first cohort is composed of the customers who made their first transaction during the first quarter of 2001, between January 1st, 2001 and March 31st, 2001. The last cohort is composed of customers who made their first transaction during the last quarter of 2002.⁵ The customers belonging to different cohorts began their relationship with the company at different market conditions. Table 1 reports the number of customers belonging to each cohort, yielding a total of $n = 11266$ customers.

5.2 Dependency between the Number of Transactions and their Average Profit

For verifying the assumption (A6), we measure the correlation, at the customer level, between the number of transactions and the average profit per transaction. Even if, at the aggregate level of all customers, the correlation

⁴We did not consider the automated pension plan transactions as stock exchange transactions. It was considered as an insurance product, to be discarded from the study.

⁵For the parameter estimation step, we discard the most extreme percentile of the customer base, i.e. customers with the 1% largest value of $x_i \times \tilde{m}_i$. These “high spending” customers are closely followed by branch agents and a global model for CLV prediction is less appropriated for them.

Table 1: Number of observations for each cohort. T01Q1 is the cohort of the first quarter of 2001, T02Q1 is the cohort of the first quarter of 2002, etc.

T01Q1	T01Q2	T01Q3	T01Q4	T02Q1	T02Q2	T02Q3	T02Q4	TOTAL
2955	1406	1114	1086	1330	1223	1287	865	11266

between the number of transactions and the average profit per transaction would be 0, or slightly positive as observed in Fader et al. [2005a], one still needs to check if this remains true at the individual level. If the number of transactions a customer i makes in a period and the profit of these transactions are correlated with a coefficient ρ_i , then, if an important group of customers has its ρ_i significantly different of 0, we will reject the independence assumption (A6).

To test hypothesis (A6), we first split the database per period of one year, then per period of two months. For each period of two months and for each customer i , the average profit per transaction and the number of transactions is observed. Let $dx_{i,s}$ be the number of transactions made by the customer i during the two months period indexed by s and $\bar{z}_{i,s} = \sum_{l \in s} z_{i,l} / dx_{i,s}$, the related average profit per transaction. Next, for each individual i and each year y , we have an estimate of the correlation between these two values in year y with $\hat{\rho}_{i,y}$ computed from 6 bimonthly observations $dx_{i,s}$ and $\bar{z}_{i,s}$. The correlation $\hat{\rho}_{i,y}$ is an imperfect estimate of ρ_i , and we test for assumption (A6) within the framework of the random effects model

$$\hat{\rho}_{i,y} = \rho_i + \epsilon_{i,y} \quad (5.1)$$

where $\rho_i \sim N(\mu_\rho, \sigma_\rho)$ and $\epsilon_{i,y} \sim N(0, \sigma_\epsilon)$ is the error term resulting from sampling and estimation error. For the independency assumption to be accepted, both μ_ρ and σ_ρ need to be equal to zero.

Using the procedure GLM of SAS 9.1.3 for random effects estimation, we compute for each customer i the estimated $\hat{\rho}_i$. Figure 1 presents the histogram of the $\hat{\rho}_i$. One can observe that there is a large heterogeneity among customers. The estimated mean of their correlations is 0.28, being significantly different from 0 ($p < 0.0001$), and an estimated σ_ρ of 0.28, again being significantly different from zero with a p smaller than .0001. We thus reject the independence assumption (A6) between the number of transactions a customer i makes and the profit these transactions yield.

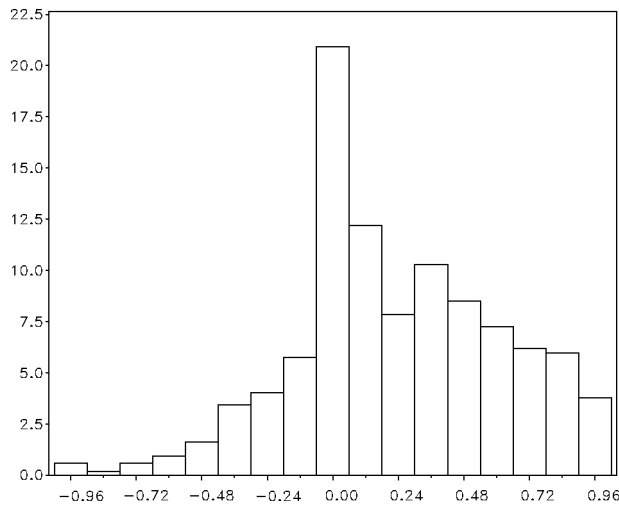


Figure 1: Histogram of the estimated correlations between the number of transactions an individual customer is making and the profit of these transactions. One estimated correlation corresponds to one customer.

5.3 Estimation of the Models

5.3.1 Pareto/NBD-Based Models

For the retail banker data set, the parameter estimation⁶ is made for all the customers who made their first transaction between January 2001 and December 2003. Then, predictions are made for the CLV of all these customers with $h = 24$ months, hence from January 2004 till December 2005. The actual data required for the computation of the true CLV for this horizon are also at our disposal, but kept out-of-sample for the model comparison.

The parameter estimates of the Pareto/NBD submodel, obtained by maximum likelihood are $\hat{r} = 0.41$, $\hat{\alpha} = 0.94$, $\hat{s} = 0.20$ and $\hat{\beta} = 1.10$. For the Gamma/Gamma submodel described in Section 3.3, we have $\hat{p} = 0.05$, $\hat{q} = 229.15$ and $\hat{\gamma} = 3.30 \times 10^5$. For equation (4.3) of the Pareto/Dependent model, the estimated coefficients are $\hat{\alpha}_1 = 1.25$, $\hat{\alpha}_2 = 0.06$, $\hat{\alpha}_3 = -0.07$ and $\hat{\alpha}_0 = -1.15$, all being highly significant. Figure 2 displays the histogram of the estimated dependency coefficients \hat{r}_i . One can see that this distribution is similar to the one of Figure 1, supporting the fact that there is non-zero correlations between the number of transactions and the related profit, which

⁶In the following, all models are implemented in Matlab 7.2. The implementation of the Pareto/NBD submodel is made on the basis of Fader et al. [2005c], available from <http://brucehardie.com/notes/008/>.

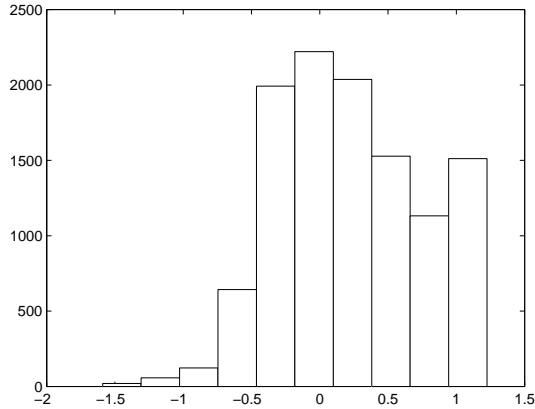


Figure 2: Histogram of the estimated dependency coefficient for the every customer in the retail banker data set as obtained from model (4.2).

motivated our new approach.

For the parameter estimation of the CDNOW data set, we obtain identical parameter estimates as Fader et al. [2005a]. For the Pareto/Dependent model of Section 4, the coefficients are $\hat{\alpha}_1 = 0.100$, $\hat{\alpha}_2 = 0.002$, $\hat{\alpha}_3 = -0.003$ and $\hat{\alpha}_0 = -0.127$. As expected, the effects here are smaller than for the retail banker data set, since a relationship between the number of transactions and the profit per transaction was not reported in Fader et al. [2005a]. However, the full regression of the dependency coefficient of equation (4.3) is overall significant.

Observing the coefficients of the regression of the dependency coefficient, one can see that the signs are the same for both data sets. The probability of activity and the cohort of a customer have a positive effect on the dependency coefficient. On the opposite, the recency of a customer has a negative effect on the dependency coefficient. We can conclude that the active and “old” customers will increase (decrease) their profit per transaction when increasing (decreasing) their number of transactions. Whereas, for the same probability of activity and cohort, the customers who made more recent purchases will decrease (increase) their profit per transaction when increasing (decreasing) their number of transactions.

5.3.2 Linear Regression

In order to compare the Pareto/NBD-based models with a simple baseline model, we also apply a linear regression formulated as

$$\text{CLV}_i = \theta w_i + \varepsilon_i, \quad (5.2)$$

where w_i is a regression vector containing an intercept, the recency t_i , the cohort T_i and the past profit yielded by the customer i being $x_i \times \tilde{m}_i$. For estimating model (5.2) we need the values of w_i , but also those of the CLV which are unknown in general. Therefore we split the sample of three years⁷ in two parts, according to the time dimension. The first part (about 22 months) is used for measuring the values of w_i , and the second (about 14 months) for measuring the actual CLV over the corresponding period. Ordinary least squares estimation yields then $\hat{\theta}$. For prediction, we compute the w_i over the complete three years time period, and do obtain the predicted values of the CLV as $\hat{\theta}w_i$. A major advantage of the Pareto/NBD-based models is that one does not need to split the estimation sample in two for the parameter estimation, since the CLV can be predicted without having a single observed value for it. Finally, note that we could not implement a linear regression for the CDNOW data set, since we only had the summary values x_i and m_i at our disposal, and not the complete transaction history.

5.4 Comparison of the Models

In this section, we compare the models previously presented on the basis of several performance measures. In Section 5.4.1 we define the accuracy measures and in Section 5.4.2 we compare the models performance at the level of the whole sample and at the level of each of the eight cohorts. We show that the Pareto/Dependent model clearly beats the Pareto/Independent model when strong correlations are present, and is still performing very well when no correlation has been reported.

5.4.1 Measures of Accuracy

As first measures of prediction accuracy, we use the *Root Mean Square Error* (RMSE) and the *Mean Absolute Error* (MAE) between the prediction of the customer lifetime value and the actual value. In order to improve the robustness to possible outliers in the data set, a trimming of 1% is applied to both the RMSE and the MAE. This process discards the largest 1% of the prediction errors. There are indeed a few customers with an extremely

⁷As before, the other two years are kept as an out-of-sample for model assessment.

large CLV, and we do not want them to dominate the analysis. The RMSE is then defined as

$$RMSE = \sqrt{\frac{1}{0.99 \times n} \sum_{i \in BP} (\widehat{CLV}_{i,h} - CLV_{i,h})^2}, \quad (5.3)$$

with n the number of observations and BP the set of the 99% best predictions. Accordingly the MAE is,

$$MAE = \frac{1}{0.99 \times n} \sum_{i \in BP} |\widehat{CLV}_{i,h} - CLV_{i,h}|. \quad (5.4)$$

As a complementary measure of prediction accuracy, we propose Spearman's correlation. Spearman's correlation is a non-parametric measure of correlation between predicted and actual values. The actual CLV and the predicted CLV are sorted by value and a rank is given to each observation. The standard correlation between these ranks is then the Spearman's correlation. We use the Spearman's correlation instead the traditional Pearson's correlation because Spearman's correlation is more robust to outliers.⁸ Moreover, this measure provides an information on the quality of the ranking of the customers based on their CLV, being a useful driver from a customer relationship perspective.

5.4.2 Results

For the retail banker data set, the securities transactions of the customers of a Belgian retail bank, the CLV is predicted for a horizon of two years on the basis of the transactions made before January first, 2004. The predicted values are then compared with the true value of the CLV yielded between January first, 2004 and the end of December 2005, discounted from the start of this period. The reader has to keep in mind that the transactions required for the computation of the actual CLV have been kept out-of-sample when estimating the models. Table 2 provides the models accuracy synthesis on the basis of the RMSE, the MAE and the Spearman's correlation. The Pareto/Dependent model outperforms the Pareto/Independent model for every accuracy measure. Moreover, even though the linear regression achieves good results, better than the Pareto/Independent model, it is still outperformed by the Pareto/Dependent approach. For this retail banker data set, the Pareto/Dependent model is clearly the best one of the three models described in this paper.

⁸Nevertheless, during our investigations, we also computed a regular correlation for each model. The ranking between the models was identical and the differences of the same order.

Table 2: Models comparison for the CLV prediction using the retail banker data set. Root Mean Square Errors, Mean Absolute Errors and Spearman’s Correlations are reported for the three considered models.

Model	RMSE	MAE	Correlation
Pareto/Independent	946.2	411.9	40.5 %
Linear Regression	892.7	340.5	47.9 %
Pareto/Dependent	843.4	324.0	51.8 %

In order to study how the models perform for different durations of transaction history, we also report the results for each cohort separately. Figure 3 shows the model accuracy as measured by the RMSE, the MAE and the Spearman’s correlation between the prediction of the customer lifetime value and the actual outcome, computed separately for each cohort. All the models show decreasing prediction accuracy when the duration of the transaction history becomes smaller.

When comparing the three models at the cohort level, several observations can be made. First, the new approach achieves a better RMSE, MAE and Spearman’s correlation over the Pareto/Independent model for almost every cohort. Secondly, the Spearman’s correlation is consistently higher for the Pareto/Dependent model than for the linear regression. When the duration of the transactions history decreases the reliability of the rank of the observation given by the linear regression also decreases, whereas it remains more stable for the Pareto/Dependent model. Moreover, a Pareto/NBD-based model has two advantages over a linear regression model. First, it does not need a splitting of the estimation sample set in two parts for the parameter estimation purpose. Secondly, it provides more information on the customer activity, since it also estimates the probability for a customer of being active. Consequently, for the retail banker data set at the cohort level, the Pareto/Dependent model is presented as the best one of the three models described in this paper.

For the sales of the CDNOW data set, Table 3 displays the accuracy of the two Pareto/NBD adaptations for the RMSE, the MAE and the Spearman’s correlation. One can see that the accuracy of the models are similar for the RMSE and the MAE measures. Nevertheless the Pareto/Dependent model has a better correlation between the ranks of the predicted and actual values. Accordingly, we claim that, even when there is only a weak dependence between the number of transactions and the average profit per transaction, the Pareto/Dependent model is still performing very well.

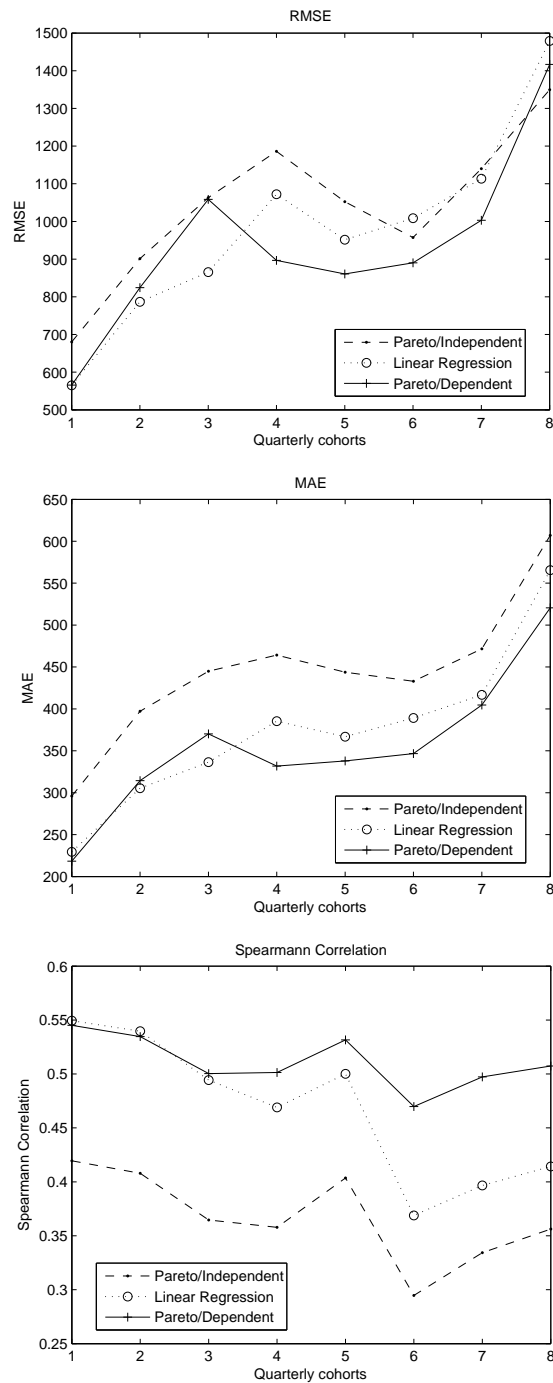


Figure 3: Root Mean Square Error (upper panel), Mean Absolute Error (middle panel) and Spearman's correlation (lower panel), computed separately for every quarterly cohort of the retail banker data set.

Table 3: Models comparison for the CLV prediction using the CDNOW data set. The Root Mean Square Errors, the Mean Absolute Error and Spearman’s Correlation are reported for the two Pareto/NBD-based models.

Model	RMSE	MAE	Correlation
Pareto/Independent	52.3	24.9	53.6 %
Pareto/Dependent	52.9	23.2	63.7 %

We can summarize the predictive accuracy results of the Pareto/Dependent approach as follows. It is not clear whether the Pareto/Dependent approach always improves a Pareto/Independent approach on the basis on a RMSE or a MAE measure of accuracy. In our application, the Pareto/Dependent clearly outperforms the Pareto/Independent approach. But, as it could be seen from the CDNOW data, when the correlation between the number of transactions and the average profit per transaction is weak, the dependency modification is not considerably improving the performance measures, but is neither deteriorating them. Nevertheless, in each of the two applications, Spearman’s correlation between the predicted and actual values of the CLV is much larger for our approach. This is a noticeable result, a model having a high Spearman’s correlation indicates a good ability to rank the customers by their CLV. This is very useful from a managerial perspective: in order to detect the most valuable customers which are to be closely followed, and for the detection of the less valuable ones, to which the company should pay less attention.

6 Conclusion

The customer lifetime value, the value of a customer based on his/her future activity, is a key metric for any business activity. The Pareto/NBD model is a suitable approach when predicting the activity of a customer in a non-contractual relationship. This paper focusses on Pareto/NBD-based models, and more particularly on the independence assumption between the number of transactions a customer makes and the average profit yielded by these transactions. We demonstrated in our empirical application that these two variables cannot be considered as independent for all customers. Predicting the future number of transactions and the future profit per transaction separately could lead to a loss in predictive performance.

A modification needs to be made in order to predict the CLV in presence of a dependence between the number of transactions a customer makes and

the profit per transaction. We proposed the Pareto/Dependent alternative, performing better in our empirical application than the Pareto/Independent model. It does not require complex adjustments. We also briefly discussed a linear regression approach, as an example of a business rule that could also be implemented. It worked quite well in the first empirical application, but it has various shortcomings as we discussed in Section 5.

Finally, in our study, only transactional data were considered. One could include socio-demographic explanatory variables for the CLV prediction. For instance, when studying the dependence between the number of transactions and the average profit per transaction as described in Section 4, socio-demographic regressors could be taken into account as well.

References

- Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, New York, 1972.
- David Bell, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz. Seven barriers to customer equity management. *Journal of Service Research*, 5(1):77–85, 2002.
- Paul D. Berger and Nada I. Nasr. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing*, 12(1):17–30, 1998.
- F. Robert Dwyer. Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing*, 11(4):6–13, 1997.
- Peter S. Fader and Bruce G. S. Hardie. A note on deriving the Pareto/NBD model and related expressions. <http://brucehardie.com/notes/009/>, 2005.
- Peter S. Fader, Bruce G. S. Hardie, and Ka Lok Lee. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4):415–430, 2005a.
- Peter S. Fader, Bruce G. S. Hardie, and Ka Lok Lee. “Counting Your Customers” the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2):275–284, 2005b.
- Peter S. Fader, Bruce G. S. Hardie, and Ka Lok Lee. A note on implementing the Pareto/NBD model in matlab. <http://brucehardie.com/notes/008/>, 2005c.

- Senil Gupta, Donald R. Lehmann, and Jennifer A. Stuart. Valuing customers. *Journal of Marketing Research*, 41(1):7–18, 2004.
- Sunil Gupta, Dominique Hanssens, Bruce Hardie, Vita Kumar, Nathaniel Lin, Nalini Ravishanker, and Sathyanarayanan Sriram. Modeling customer lifetime value. *Journal of Service Research*, 9(2):139–155, 2006.
- Dipak Jain and Siddhartha S. Singh. Customer lifetime value research in marketing: A review and future directions. *Journal of Interactive Marketing*, 16(2):34–46, 2002.
- Edward C. Malthouse and Robert C. Blattberg. Can we predict customer lifetime value? *Journal of Interactive Marketing*, 19(1):2–16, 2005.
- Francis J. Mulhern. Customer profitability analysis: Measurement, concentrations, and research directions. *Journal of Interactive Marketing*, 13(1):25–40, 1999.
- Rakesh Niraj, Mahendra Gupta, and Chakravarthi Narasimhan. Customer profitability in a supply chain. *Journal of Marketing*, 65(3):1–16, 2001.
- Phillip E. Pfeifer, Mark R. Haskins, and Robert M. Conroy. Customer lifetime value, customer profitability, and the treatment of acquisition spending. *Journal of Managerial Issues*, 17(1):11–25, 2005.
- Werner J. Reinartz and Vita Kumar. On the profitability of long-life customers in a non contractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, 64(4):17–35, 2000.
- Werner J. Reinartz and Vita Kumar. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67(1):77–99, 2003.
- Roland T Rust, Katherine N. Lemon, and Valarie A. Zeithaml. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing*, 68(1):109–127, 2004.
- David C. Schmittlein and Robert A. Peterson. Customer base analysis: An industrial purchase process application. *Marketing Science*, 13(1), 1994.
- David C. Schmittlein, Donald G. Morrison, and Richard Colombo. Counting your customers: who are they and what will they do next? *Management Science*, 33(1), 1987.

Rajkumar Venkatesan and Vita Kumar. A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4):106–125, 2004.