



FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 - 264.34.61
Fax. : 32 - (0)9 - 264.35.92

WORKING PAPER

**Benefits of Quantile Regression for the Analysis of
Customer Lifetime Value in a Contractual Setting:
An Application in Financial Services**

Dries F. Benoit¹

Dirk Van den Poel²

January 2009

2009/551

¹ PhD Candidate, Ghent University

² Corresponding author: Prof. Dr. Dirk Van den Poel, Professor of Marketing Modeling/analytical Customer Relationship Management, Faculty of Economics and Business Administration, dirk.vandenpoel@ugent.be; more papers about customer relationship management can be obtained from the website: www.crm.UGent.be

Benefits of Quantile Regression for the Analysis of Customer Lifetime Value in a Contractual Setting: An Application in Financial Services

Dries F. Benoit* & Dirk Van den Poel*

*Ghent University, Faculty of Economics and Business Administration, Department of Marketing,
Tweakerkenstraat 2, B-9000 Ghent, Belgium.

Abstract. The move towards a customer-centred approach to marketing, coupled with the increasing availability of customer transaction data, has led to an interest in understanding and estimating customer lifetime value (CLV). Several authors point out that, when evaluating customer profitability, profitable customers are rare compared to the unprofitable ones. In spite of this, most authors fail to recognize the implications of these skewed distributions on the performance of models they use. In this study, we propose analyzing CLV by means of Quantile Regression. In a financial services application, we show that this technique provides management more in-depth insights into the effects of the covariates that are missed with Linear Regression. Moreover, we show that in the common situation where interest is in a top-customer segment, Quantile Regression outperforms Linear Regression. The method also has the ability of constructing prediction intervals. Combining the CLV point estimate with the prediction intervals leads to a new segmentation scheme that is the first to account for uncertainty in the predictions. This segmentation is ideally suited for managing the portfolio of customers.

Keywords: customer relationship management (CRM), database marketing, customer segmentation, quantile regression, prediction interval, customer lifetime value

Corresponding author: Dirk Van den Poel (Dirk.VandenPoel@UGent.be);

Tel. +32 9 264 89 80; Fax. + 32 9 264 42 79;

Website research: <http://www.crm.UGent.be>

Website teaching: <http://www.mma.UGent.be>

1. Introduction

Over the past decade, Customer Relationship Management (CRM) has become a leading strategy in highly competitive business environments. Companies increasingly derive revenue from the creation and enhancement of long-term relationships with their customers [5]. This move towards a customer-centric approach to marketing, coupled with the increasing availability of customer-transaction data, has led to an interest in estimating and understanding Customer Lifetime Value (CLV). CLV is viewed as the present value of the future cash flows associated with a customer [29]. Knowing the CLV of individual customers enables the decision maker to improve the customer segmentation and marketing resource allocation efforts [24, 20] and this in turn will lead to higher retention rates and profits for the firm [15].

Donkers et al. [6] give a detailed overview and comparison of the wide range of different approaches that have been used for CLV modeling. From their outline it is clear that regression-type models are often used in this context (e.g. Linear Regression model [10, 25, 14, 38]; Probit model [3]; multivariate Probit model [6]; multivariate Logit model [30]). Several authors [19, 8, 11] point out that, when evaluating customer profitability, profitable customers are rare compared to the unprofitable ones. Gupta et al. [13] remark that the regression-type models easily break down when applied to settings where the behavior of interest is rare. In spite of this, most authors fail to recognize the implications of this skewed distribution for the models they use. Note that in this study, the focus is on the situation where the buyer-seller relationship is governed by a contract. The Pareto/NBD model, which is state-of-the-art in a non-contractual setting [10, 39], is inappropriate in this contractual context [35] and is therefore excluded from the current analysis.

In this paper, we propose to analyze CLV by means of Quantile Regression. Quantile Regression [22,21] is a method for fitting a regression line through the conditional quantiles of a distribution. Therefore, the method is less influenced by long-tailed, skewed distributions that are typical in CLV modeling. Moreover, the manager's interest is often not in the large group of unprofitable customers,

but in the smaller group of more lucrative customers. In this case, mean regression gives only very limited information and it is worthwhile considering the more extreme quantiles of the CLV distribution. Consequently, explicit investigation of the effects of the covariates via Quantile Regression can provide a more nuanced view of the stochastic relationship between the covariates and the dependent variable [23]. Therefore Quantile Regression results in a more informative empirical analysis. Furthermore, Quantile Regression produces prediction intervals that give insight into the uncertainty about a CLV prediction. This information is important for the decision maker when quantifying the risk associated with any given customer [13].

This study contributes to the existing customer lifetime value literature by investigating the usefulness of the Quantile Regression approach in the financial services industry. First, we demonstrate that by using Quantile Regression, we provide management with a more detailed insight into the complex relationship of the CLV drivers than mean regression does. Secondly, we show that Quantile Regression outperforms the often used mean regression as predictive tool in the context of CLV modeling. Thirdly, we provide our CLV forecasts with prediction intervals or even with the entire predictive distribution. This makes the task of customer risk assessment more straightforward for the manager. A new segmentation scheme is proposed, based on CLV forecast and the related prediction interval. This method of customer segmentation distinguishes itself from other CLV schemes by explicitly taking uncertainty into account.

The structure of this paper is as follows. In Section 2, we review the previous studies related to customer value. It illustrates the limitations of existing studies and sets the stage for this paper. In Section 3, we describe our methodology for the analysis and prediction of customer lifetime value. In Section 4, we present an observational application of this methodology in the financial services industry. We also discuss the market segmentation and managerial implications for this application. Finally, Section 5 concludes with remarks on the limitations of this study and future research directions.

2. Related works

Customer Lifetime Value has been studied under the name of LTV, Customer Value, Customer Equity and Customer Profitability. The concept is defined as the sum of the revenues gained from company's customers over the lifetime of transactions after deduction of the total cost of attracting, selling and servicing customers, taking into account the time value of money [16]. The basic formula for calculating CLV for customer i at time t for a finite time horizon T [1] is:

$$CLV_{i,t} = \sum_{\tau=0}^T \frac{\text{Profit}_{i,t+\tau}}{(1+d)^\tau}, \quad (1)$$

Where d is a pre-determined discount rate. In multi-service industries, $\text{Profit}_{i,t}$ is defined as:

$$\text{Profit}_{i,t} = \sum_{j=1}^J \text{Serv}_{ij,t} * \text{Usage}_{ij,t} * \text{Margin}_{ij,t} \quad (2)$$

Here J is the number of different services sold, $\text{Serv}_{ij,t}$ is a dummy indicating whether customer i purchases service j at time t , $\text{Usage}_{ij,t}$ is the amount of that service purchased and $\text{Margin}_{ij,t}$ is the average profit margin for service j .

Theoretically, CLV models should estimate the value of a customer over the entire customer's lifetime. However, in practice most researchers use a finite time horizon of three or four years (e.g. [6, 34]). Three to four years is a good estimate for the horizon over which the current business environment would not substantially change and even then, there is significant uncertainty in predicting customer behavior [37]. Moreover, some research considers an even shorter time horizon [16].

CLV has been analyzed in a substantial number of different domains, varying from econometric models to computer science techniques. However, the key questions are usually very similar: "*What are the drivers of CLV?*", "*Which customers are the future most valuable ones?*", "*How to address the top customers?*", etc. Several authors give an overview of the variety of modeling procedures that were used in search for answers to the key questions [27, 13, 6, 1, 36]. In general, one can distinguish two broad classes of models in the current contractual setting. First, a large group of models focuses

on the choices customers face when buying an additional service or product. A customer's lifetime value is then seen as the sum of the distinct contributions per service or product. This approach is appealing because of the natural way in which the CLV prediction is built up. In a first stage, an estimation is made on the probability of a customer buying a given product or service. The second stage is then to combine these probabilities with the margins associated with the product or service into an aggregate prediction of a customer's lifetime value. This approach also has the advantage of providing more insight into the factors that drive customer value. The main drawbacks are the amount of modeling required and the often poorer predictions. Examples of this approach are found in Venkatesan and Kumar [36] and Hwang et al. [16]. The second large group of models does not follow the two stage method, but focuses directly on relationship length and total profits. Since the individual-level choice modeling is left aside, the process of producing CLV estimates is much more straightforward and prediction accuracy is higher [38]. As such, this approach turns the disadvantages of the first approach into benefits. However, due to aggregation, insight into the factors that drive consumer profitability is limited compared to the choice-based approach. Examples of CLV research following this direct approach are found in Malthouse & Blattberg [25] and in Hansotia and Rukstales [14].

Given that one of the key issues when decision makers use the CLV metric is whether the firm can provide an adequate prediction of the CLV of each customer in the database [25, 36], it is clear that the predictive accuracy of the CLV is of primordial importance. Furthermore, these predictions are often used as guidelines for investments in segments of customers [43]. However, the previously used regression techniques are often not ideally suited for the purpose of modeling customer lifetime value. When evaluating customer profitability, marketers are often reminded of the 80/20 rule (80% of the profits are produced by top 20% of profitable customers and 80% of the costs are produced by top 20% of unprofitable customers) [8, 11]. This finding has important implications for both the two-stage approach as well as for the approach that models CLV directly. For researchers using the two-step CLV approach, the problem arises when modeling the choice problem. Since the largest group of customers buys no or only a very limited amount of products or services and only a small group of

customers buys many products or services, the researcher should be aware of the fact that he or she is modeling rare events. In this rare-event situation, it is known that parametric choice models easily break down [13]. The other approach, where the researcher focuses directly on the relationship length and total profits, leaves aside the individual-level choice modeling step. However, the problem of rare events can not be totally avoided. This is because the underlying process (the 80/20 rule) results in a lifetime value variable that tends to have a strong non-normal distribution and the usual assumption of homoscedasticity is hard to maintain [9, 25]. In contrast, the proposed Quantile Regression technique does not suffer from these particularities of the CLV variable [4]. Quantile Regression can be seen as part of the second approach since it models CLV directly. This direct approach often leads to high predictive performance, but Quantile Regression also provides the manager with insights in the covariates that are totally missed with other methods from the direct approach (e.g. Linear Regression). Thus, Quantile Regression combines somewhat the best of the two approaches.

A large group of researchers have recommended CLV as a metric for selecting customers and designing marketing programs (e.g. [31, 34, 36, 19, 38]). One of the key reasons for this is the finding that customers selected on the basis of CLV generate more profits than customers selected on the basis of other metrics such as socio-demographics [31, 36]. Several segmentation strategies based on CLV are proposed, but they generally can be classified into three categories [19]: (i) segmentation by using only CLV values (e.g. the customer pyramid [43]), (ii) segmentation by considering both CLV values and other information (e.g. socio-demographic information, transaction history, etc. [19]) and (iii) segmentation by using only CLV components (e.g. current value, potential value, loyalty, etc. [16]). A major drawback all the segmentation proposals have in common is that uncertainty about the CLV prediction is not taken into account. When talking about CLV, researchers tend to consider the expected value of CLV, which may not always be appropriate. As such, acting on this information, the marketing manager may focus on acquiring high-CLV customers without considering the possible high risk or uncertainty that is associated with the customer. Since the financial market expects the firm to have a portfolio of customers that comprises a mix of low- and high-risk customers, this way of decision making may be suboptimal for the firm [13]. As displayed in Figure 1, we therefore

propose a segmentation matrix based on two important aspects of future value, notably the CLV prediction and the uncertainty around the prediction. Quantile Regression is ideally suited for this, since it naturally provides a prediction with a measure of uncertainty, i.e. the prediction interval, without relying on a theoretical normal distribution. Using Quantile Regression with the resulting segmentation scheme explicitly forces the manager to take uncertainty into account and avoids disproportionate investment in customers with high risk without considering it. In the application in Section 4.5, we will show the relevance of this segmentation for the decision maker.

INSERT FIGURE 1 ABOUT HERE

From this outline it should be clear that, even through the research concerning CLV is quite elaborate, there still remain quite some challenges for improvement. Quantile Regression improves the existing approach to CLV on several aspects. Quantile Regression is a useful and powerful forecasting tool, also providing the decision maker with unique insights in the complex relationship between the covariates and the CLV variable. Moreover this technique is able to provide prediction intervals that inform the manager about the uncertainty that is associated with the CLV prediction of a customer. The CLV forecast together with the prediction interval can then be used as segmentation scheme.

3. Methodology

3.1 Quantile Regression vs. Mean Regression

The classical theory of linear models focuses on the conditional mean function, the function that describes how the mean of y changes with the vector of covariates x . This least-squares approach assumes that the error has exactly the same distribution whatever values of x may be taken. The components of the vector of x are expected to affect only the location of the conditional distribution of y , not its scale or any other aspect of its distributional shape. If this is the case, a researcher may be fully satisfied with an estimated model of the conditional mean function. However, in many situations it is interesting to go beyond this classical least-squares regression approach.

Quantile Regression [22, 21] extends the well-known mean regression model to conditional quantiles of the response variable, such as the median. This approach provides a more nuanced view of the relationship of the dependent variable and the covariates, since it allows the user to examine the relationship between a set of covariates and the different parts of the distribution of the response variable. Dotson et al. [7] provide a simple example of the concept. The data for Figure 2 were generated according to the heteroscedastic regression model in Equation 3. Each fitted line represents the estimated relationship between the independent variable, x , and a particular quantile of the dependent variable, y . Since the error variance of y is positively correlated with x , slope coefficients differ across quantiles.

$$y = 2x + \varepsilon; \quad \varepsilon = N(0, 0.6x) \quad (3)$$

INSERT FIG. 2 ABOUT HERE

Consider the standard regression model where y and x are both continuous variables:

$$y_i = \mu(x_i) + \varepsilon_i \quad (4)$$

If we assume that $E(\varepsilon|x) = 0$, then $\mu(x_i)$ is a conditional mean function, while if $\text{Med}(\varepsilon|x) = 0$, then $\mu(x_i)$ is a conditional median function. Since we assume that the relation between y and x is linear, we obtain a conditional expectation model:

$$E(y_i | x_i) = x_i' \beta \quad (5)$$

or a linear conditional median model:

$$\text{Med}(y_i | x_i) = x_i' \beta \quad (6)$$

depending on what restriction we have put on the error term.

In the conditional expectation model (Equation 5), we can find the regression coefficients by solving:

$$\arg \min_{\beta \in \mathbb{R}} \sum_{i=1}^n (y_i - x_i' \beta)^2 \quad (7)$$

One could say that we try to find a value for β that minimizes the sum of squared errors. In median regression, we proceed in exactly the same way, but here, we try to find an estimate of β that minimizes the sum of the absolute deviations:

$$\arg \min_{\beta \in \mathfrak{R}} \sum_{i=1}^n |y_i - x_i' \beta| \quad (8)$$

Quantile Regression proceeds by extending the median case to all other quantiles of interest. Recall that the general τ^{th} sample quantile $\xi(\tau)$, $0 < \tau < 1$, can be formulated as the minimizer:

$$\xi(\tau) = \arg \min_{\xi \in \mathfrak{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \xi) \quad (9)$$

where $\rho_{\tau}(z) = z(\tau - I(z > 0))$ and where $I(\cdot)$ denotes the indicator function. The loss function ρ_{τ} assigns a weight of τ to positive residuals and a weight of $(1 - \tau)$ to negative residuals. Using this loss function, the linear conditional quantile function extends the τ^{th} sample quantile $\xi(\tau)$ to the regression setting in the same way as the linear conditional mean or median function.

$$\beta_{\text{hat}}(\tau) = \arg \min_{\beta \in \mathfrak{R}} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta) \quad (10)$$

for any quantile $\tau \in (0,1)$. The quantile $\beta_{\text{hat}}(\tau)$ is called the τ^{th} regression quantile. Note that the case where τ equals 0.5, which minimizes the sum of absolute residuals, corresponds to the median regression.

It is possible to regard the problem using a Bayesian approach. The use of Bayesian inference for generalized linear models is quite standard these days [2]. Two obvious advantages of the Bayesian approach over classical inference are that the Bayesian approach may lead to exact and full inference conditional on the observed data, as opposed to the asymptotic inference of the classical approach, and that Bayesian inference deals in a better way with parameter uncertainty [40]. Yu and Moyeed [41] show that the minimization problem in Equation 10 is equivalent to maximizing a likelihood function composed of independently distributed Asymmetric Laplace (AL) densities. Their proof lays the framework for development of a Bayesian approach to quantile regression. Yu and Zhang [42] propose a three-parameter AL that shows to have ideal properties in the context of quantile regression:

$$f_p(y | \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\} \quad (11)$$

where

$$\rho_\tau(y) = y(\tau - I(y < 0)) \quad (12)$$

Equation 12 is identical to the loss function in the programming problem in Equation 9. Thus, minimizing Equation 10 is equivalent to maximizing a regression likelihood using AL errors with $\mu = \beta'X$. Bayesian implementation of quantile regression begins by forming a likelihood comprised of independent AL densities with $\mu = \beta'X$, specifying the quantile of interest, τ , and placing priors on the model parameters β and σ . Posterior inference about model parameters then follows conventional Bayesian procedures (see for example [32]).

3.2 Predictive performance measures

The predictive performance of the models is assessed with respect to two different tasks. First, predicting the absolute level of individual CLV and second, predicting the ordering of customers based on CLV.

Correct predictions for the level of CLV are relevant when a company wants to target customers with a CLV above a given level. This is the case in situations where the expected CLV has to be above a certain level in order to have a profitable customer. Often used measures as (root) MSE or the MAD criteria are not appropriate to make the current comparison, since the regression techniques optimize for one of the two criteria. Linear regression will have the best predictive performance when the root MSE criterion is used, while quantile regression (with $\tau=0.5$) excels when the MAD criterion is considered. To analyze how well each model predicts the level of CLV for each individual, we therefore use the hit-rate criterion proposed by Donkers et al. [6]. For this hit rate, we categorize all customers based on their true CLV into four equal-sized groups with increasing levels of CLV. The hit rate is then computed as the percentage of customers whose predicted CLV falls into the same category as their true CLV.

However, more often the management will be interested in targeting their most profitable customers without being interested in the precise level of CLV. In the latter case, it is more appropriate to use an ordering-based measure of predictive performance. Since CLV is more often used as a segmentation device than as a tool to manage profitability of marketing activities at the individual level [43], the ordering-based measure will be more relevant. The predictive performance with respect to the ordering of the customers is also evaluated with a hit-rate criterion. Several studies [25, 6] use it as a measure of predictive performance with respect to the ordering of the customers. In contrast to the hit rate for levels, this hit rate does not consider the level of CLV, but only the ordering of the customers with respect to CLV. The ordering-based hit rate measures what percentage of customers with a true CLV in the top- x percent have a predicted CLV that is also in the top- x percent based on predicted CLV.

3.3 Prediction intervals

How reliable is a CLV prediction for a new customer? For some households, it might be possible to pinpoint future CLV to a higher accuracy than for other families. With standard mean regression techniques, a single point estimate is returned for each new instance. However, this point estimate does not contain information about the dispersion of observations around the predictive value. Based on asymptotic normality, prediction intervals can be obtained, but Quantile Regression offers a principled way of judging the reliability of prediction [26]. Gupta et al. [13] emphasize that modeling the expected value of CLV is not always appropriate. Therefore, we also must consider the variance so that we can move towards quantifying the risk associated with any given customer. Quantile regression can be used to build such prediction intervals. For example, a 90% prediction interval for the value of Y is given by

$$I(x) = [Q_{.05}(x), Q_{.95}(x)] \quad (13)$$

That is, a new observation of Y , for $X = x$, is with high probability (e.g. 90%) in the interval $I(x)$. The width of this prediction interval can vary greatly with x [26].

4. A financial services application

In this section, we present the application of our methodology to a real-life setting. This section is organised as follows: In a first part, detailed information about the financial services setting is given. The second part provides the results concerning the insights Quantile Regression provides into the complex relationship between the covariates. The third part focuses on the predictive performance of the model, while the fourth part covers the prediction intervals. Finally, the last part shows the possibilities of the proposed segmentation scheme.

4.1 Data

The data are provided by a large European financial services company. The database contains both transactional and socio-demographic information about the customers. For the current analysis, we work at the household level. Households are studied, instead of individuals, because in this market the household is the principal decision-making unit [12, 28]. A household is defined as all customers living on the same address. We used a total of 22,665 families that had at least one active product on December 31, 2003. The CLV is then computed, using Equations 1 and 2, for the four subsequent years. In line with the business practice of the company, we use margins that are not consumer specific and the usage levels are set to their expected values. Also note that the margin is calculated taking into account the defection rates of the customers buying service j . The entire dataset is divided into training set and validation set at the ratio 60/40, respectively. Performance measures are reported only on the validation set. For the independent variables, we include past behavioral data as well as socio-demographic data. Table 1 gives an overview of the different variables that are used in the analysis. Most of these variables have proven to be good predictors of lifetime value in previous studies [33, 9, 19]. Other variables are chosen because they are of special interest for the company. Note that all continuous variables are standardized around their mean value. This is done for numerical stability, but also because of the confidential nature of the data.

INSERT TABLE 1 ABOUT HERE

As pointed out in Section 2, the dependent variable is often strongly right skewed in lifetime value modeling. Malthouse and Blattberg [25] suggest a variance stabilizing transformation of the CLV variable when using Linear Regression. According to their [25] practice, the logarithmic transformation is used and a constant is added to every value to avoid taking the logarithm of a negative value or zero. By doing so, a more fair comparison is made between the different modeling techniques. Note that in the current application this transformed variable is not only used for Linear Regression, but also for Quantile Regression.

4.2 Lifetime value drivers

Although much research on customer lifetime value has employed conventional least-squares regression methods, it has been recognized that the resulting estimates of various effects on the conditional mean of CLV were not necessarily indicative of the size and nature of these effects on the lower or upper tail of the CLV distribution [13]. A more complete picture of covariate effects can be provided by estimating a family of conditional quantile functions.

Figure 3 presents a concise summary of the quantile regression results for this application. Each plot depicts one variable in the Quantile Regression model, $\{\hat{\beta}_j(\tau) : j = 1, \dots, 11\}$. Note that the plots in Figure 3 are obtained using Bayesian estimation with vague priors on the unknown model parameters. The plotted point estimates and the credible intervals are the expectation, $Q_{.025}$ percentile and $Q_{.975}$ percentile obtained from the marginal posterior distribution of the different parameters. The solid line with filled dots represents the point estimates of the regression coefficients for the different quantiles, $\{\tau_q : q = 0.05, \dots, 0.95, \text{ by } 0.5\}$. The shaded area depicting a 95% pointwise credible band is obtained from the marginal posterior distribution of the different parameters. Superimposed on the plot is a dashed line representing the ordinary least-squares estimate of the mean effect, with two dotted lines representing a 95% confidence interval.

INSERT FIGURE 3 ABOUT HERE

At first glance one can see that for this application it might be worthwhile to split up the effects of the covariates in three distinct parts when the results are discussed, τ in $[0.05,0.3]$, τ in $[0.3,0.7]$, τ in $[0.7,0.95]$. For most variables, the effect on the lower conditional quantiles is very small and often not significant. When considering the effects on the more central conditional quantiles, one can see that the effect of the variables is often quite constant over the interval (e.g. `Social_Class_Score`, `FTHB_gez`, `nbr_insur`, `lor` and `recency`). However, this is not the case for all variables (e.g. `nbr_FamInd`, `freq`, and `agent`). For the upper quantiles of the conditional CLV distribution, the effect of most variables shows to increase or decrease a lot, compared to the lower quantiles. For all but one variable (i.e. `Age_max`) the effect of the covariate is most pronounced in the upper quantiles of the conditional CLV distribution.

In the first panel of Figure 3, the intercept of the model may be interpreted as the estimated conditional quantile function of the CLV distribution of a household that buys through an agent, not being part of the company defined segment, with mean maximum age of the household members, mean number of household members, mean number of foreign household members, mean social class score, mean purchase frequency, mean number of assurances, mean length of relation and mean recency.

The maximum age of the household members clearly has a negative effect on the CLV measure. For this variable, one can see that the effect in the upper quantiles is close to the mean (OLS) effect. However, the estimated conditional median is more pronounced than the conditional mean effect. Verhoef & Donkers [38] also show the importance of age in a CLV model.

Not surprisingly, the number of household members has a positive impact on the expected lifetime value. The estimated conditional quantile function reveals that the effect of the covariate is getting more extreme for the higher quantiles. For this variable, the effect on the conditional mean is quite similar to the effect on the conditional median of CLV distribution.

The number of foreign household members exerts a negative effect on the lifetime value estimate. In contrast to some of the other effects, the effect of this covariate is quite stable over the entire distribution, as indicated by the fact that the quantile function lies almost entirely between the 95% confidence intervals of the OLS estimate. Moreover, for the extreme quantiles, the parameter is not significantly different from zero. The added value of quantile regression estimates is rather limited for this variable.

The least-squares estimate for the social class score suggests that the effect is positive. However, the confidence bands around the OLS estimate show that this effect is not different from zero ($p > 0.05$). Here, Quantile Regression gives a more detailed insight in the effect of the covariate. The plot reveals that the covariate has no significant effect on the conditional CLV distribution for the lower and middle quantiles. But for the higher quantiles this effect is clearly different from zero. Ample research supports the finding that higher social class families are more profitable than lower social class families [31]. But here we can add the insight that this effect becomes more and more pronounced when considering higher quantiles of the CLV distribution.

The segmenting variable (FTHB_gez) is a striking example of how misleading the OLS estimates can be. Again, the OLS estimate suggests that the effect of the variable is not different from zero. But when considering the Quantile Regression results, a different picture becomes clear. The conditional quantile function of CLV reveals a positive effect for the middle quantiles. However, in the upper quantiles of the CLV distribution, the covariate has a negative effect. Recall that this variable is a dummy indicating whether the household is a young family with a first real-estate project. It could be argued that families being in that specific stage of their family life cycle acquire a rather modest amount of financial products. On the other hand, those families are often financially immature and therefore they are often buying products from product categories that are less profitable for the financial services company [17], which is indicated by the negative relationship in the upper quantiles of the CLV distribution. This is in line with the findings of [43] that young professionals who purchase homes at the upper end should be tagged as potential top customers.

Whether the household interacts with the company through an agent or not (agent), shows to exert a different effect for the different quantiles. However, the mean and median effects are very similar and in this case, the OLS estimate produces a satisfying summary of the conditional quantile distribution.

From literature one could expect a positive relationship for the frequency variable [31, 25]. Note that the OLS estimate is very similar to the median conditional quantile estimate. However, for the higher quantiles, the OLS estimate is strongly underestimating the effect of the variable.

The number of insurance policies and the length of relationship exert a rather similar effect on the conditional CLV distribution. For the lower quantiles the effect is not significant, for the middle quantiles the effect remains rather constant, but for the higher quantiles the effect of the variables increases fast. The variable recency seems to exert a similar effect, but here in the opposite direction. For these three variables, the OLS estimate is a bit more extreme than the effect in the middle quantiles. For the extreme quantiles, the OLS estimate is again a substantial under- or overestimation of the real effect.

The parameters in Figure 3 clearly show that quantile regression results offer a much richer, more focused view of the application than could be achieved by looking exclusively at conditional mean models. In particular, it provides a way to explore sources of heterogeneity in the response that are associated with the covariates.

4.3 Predictive results

The added value of Quantile Regression is not limited to the richer insights into the effects of the covariates. In this part we focus on the out-of-sample predictive performance of the quantile regression models. In a first step, we will focus first on predicting the level of individual CLV. Secondly, our interest is on predicting the ordering of the customers based on CLV.

We start by investigating the performance of the different models for predicting the absolute level of CLV. Therefore, we set $\tau=0.5$ and the quantile regression reduces to median regression. Table 2 reports the results for the models we compare. The naïve model is a model without explanatory variables and always predicts the mean.

INSERT TABLE 2 ABOUT HERE

Table 2 shows the results for the hit-rate criterion explained in Section 3.3. Several Chi-square tests indicated significant ($p<0.01$) differences between all models. This means that Quantile Regression model with $\tau = 0.5$ performs better than the models without explanatory variables and better than the mean regression model in terms of absolute predictions. Thus, when management wants to target a group of customers with an expected CLV above a specific threshold level, the quantile regression approach is more appropriate. In the case of lifetime value, however, the managers' focus is often not on the absolute CLV levels, but rather on the ordering of the customer base. As explained in Section 3.3 the main reason for this is the popularity of CLV as basis for customer segmentation. Therefore, it might be more relevant to evaluate the predictive performance of the ordering of the customers.

INSERT FIGURE 4 ABOUT HERE

Figure 4 gives the results for the different models regarding the predictive performance of the ordering of the customers. When the ordering of the customers is of primary interest, Quantile Regression takes advantage of its properties of handling asymmetrical errors. When, for example, a prediction for the top 5% is required, one applies the regression estimates of the Quantile Regression where $\tau = 0.95$. The evaluation criterion used is justified and explained in Section 3.3. Note that a naïve model would randomly assign the customers to the top group. When, for example, one tries to predict the top-30% customers, the naïve model randomly assigns 30% of the customers to that condition. Note that the performance of this model is so low that it falls outside the plotting area of Figure 4. The horizontal

axis represents the top- x percent one tries to predict. The vertical axis is then the performance of the models based on the ordered hit-rate criterion. A Chi-square test is performed to check the significance of the difference in performance; when significant ($p < 0.05$) the area between the two curves is shaded.

As shown in Figure 4, the advantage of using Quantile Regression is most pronounced when the interest is in the high-end customers and wears off when larger top groups are considered. Also note that the performance of the mean regression deteriorates at a higher rate than the Quantile Regression performance when the focus is on decreasing top segments. Knowing that those top groups are often of central interest for the decision makers, the use of Quantile Regression in this context makes a real difference.

4.4 Prediction intervals

As discussed in Section 3.4, Quantile Regression offers the possibility of constructing prediction intervals. For each new data point x_i , a prediction interval of the form used in Equation 13 gives a range that covers the new observation of the response variable Y with high probability. Figure 5 shows some graphical results for the current application.

INSERT FIGURE 5 ABOUT HERE

Figure 5 plots each observation together with its prediction interval. For better visualisation, observations are ordered according to the length of their corresponding prediction interval and the mean of the upper and lower end of the prediction interval is subtracted from all observations and prediction intervals. As can be seen from the graph, the prediction intervals vary in length, some being much shorter than others. Additionally, the percentages of observations that lie above the upper end (below the lower end) of their respective prediction intervals are indicated in the upper (lower) left corner. The latter figure shows that the prediction intervals of Quantile Regression work very well, knowing that theoretically 5% of all observations should be above and below their 90% prediction interval. Using quantile regression it is possible to give a range in which, with high probability, each

observation is going to be. This gives an idea to the decision maker about how reliable a given CLV prediction is and assesses the risk associated to that particular customer. One can also look at the prediction interval as arising from a different viewpoint of the researcher. The median point estimate is then seen as a neutral estimate of what level of CLV is expected conditional on the covariates. Within this viewpoint, the upper bound of the prediction interval (e.g. $Q_{.90}$) is a very optimistic prediction, while the lower bound of the interval (e.g. $Q_{.10}$) is a very pessimistic prediction of future CLV conditional on the predictors.

4.5 Segmentation scheme

So far, we have discussed the estimation and prediction results for CLV. The remaining question is whether these results can be used to construct a useful segmentation of the customer base. Section 2 showed that previous research has proposed several segmentation schemes. However, a major shortcoming is that all those schemes ignore the notion of uncertainty. Quantile Regression offers a natural approach for predictive uncertainty.

INSERT FIGURE 6 ABOUT HERE

Figure 6 plots the point estimates of the CLV predictions using $\tau=0.5$ versus the uncertainty associated with the prediction. First, one immediately notices that the bulk of the data is situated in the lower left corner of the graph, showing that the company has lots of customers whose expected median CLV is rather low and also the uncertainty around this prediction is low. In Section 2 we named this group segment I. At first thought, this group may be regarded as unimportant or even unwanted. However, this segment can be important for the creation of economies of scale [18] or in industries where network effects exist [20]. Second, in the lower right corner the high profit – low uncertainty customers are situated (segment II). There is no doubt that this is the most valuable segment for the company. As is seen from the graph, the analysis supports the idea that when the expected median CLV increases, the uncertainty about this prediction is increasing too. Next, in the upper right corner, we find the households with a high expected median CLV, but this expectation is rather uncertain

(segment III). It is unburdened that this segment should not be treated equally as the high profit – low uncertainty segment, as previous research would have done. The decision maker should be aware of the fact that investing in this group has a more uncertain outcome than investments in segment II. Finally, the fourth segment contains the families with a low expected median CLV, but this prediction is rather uncertain. This implies that this segment might deserve some more attention from the company compared to segment I. For both segments one expects low profitability, but for the customers in segment IV it might be more profitable for the company to give them the benefit of the doubt.

We do not want to explicitly set up bounds that distinguish between high and low CLV customers. Nor do we want to determine the exact distinction between certain and uncertain predictions. With regard to segmentation based on lifetime value, a median split is often used in previous research [38]. This is the split we used for Figure 6. One could also argue in favor of the 80/20 rule [8, 11]. However, more important than providing an exact cut-off is that the information from the prediction intervals enables management to value the portfolio of customers and develop rules that guide the marketing manager to undertake actions that maximize the expected value of the portfolio rather than the value of the next-acquired customer. And this is what is of central importance when customers are managed [13]. The segmentation scheme helps the decision maker to consider whether it is better to acquire a few large customers (who may be risky) or a large number of small customers.

5. Conclusion

In this paper, we show the benefits of using Quantile Regression for modeling customer lifetime value. To our knowledge, this study is the first to make use of Quantile Regression in the context of CLV modeling. Even in the marketing domain as a whole, this technique is rarely used (i.e. [7]). Ample research showed that in the context of CLV modeling the dependent variable is often skewed. This is an important, but often neglected feature of the modeling task and Quantile Regression is an

appropriate technique in such a setting. We showed that Quantile Regression provides insights into the effects of the covariates on the conditional CLV distribution that may be missed by traditional least-squares estimates. For some covariates the difference was very extreme: The effect they exerted had an opposite sign for different parts of the conditional life time value distribution, while the OLS estimate showed no significant effect. It is clear that interpretation of the Quantile Regression estimates has led to a more detailed and thorough insight into the effects of the drivers of CLV.

Previous studies had found mean regression as the best performing modeling technique when the focus is on predictive purposes. However, for the current financial services application, Quantile Regression outperformed the traditional regression on both absolute predictive performance and the ordering-based predictive performance. The most striking difference was found in situations where one wants to predict a given top percentage of customers. Managers face this kind of problem setting very often, for example when they want to target the top- x percent of most profitable customers for some marketing action. The smaller the top segment of interest, the better the predictive performance of quantile regression compared to least-squares solutions.

The properties of quantile regression also make it possible to construct prediction intervals around a CLV point estimate. We showed that these intervals were not of equal length for every household in the analysis. The prediction intervals give insight into the uncertainty associated with a CLV prediction at the individual customer level. Therefore, the prediction intervals can be seen as measures of the risk linked to the individual. Alternatively, the different quantiles can be thought of as representing a different level of optimism/pessimism regarding the future lifetime value of the customer. To our knowledge, this is the first study that fully accounts for uncertainty regarding lifetime value predictions.

CLV is often used for segmentation purposes. We believe that a segmentation scheme based on the combination of a lifetime value point estimate together with the size of the prediction interval results in a valuable marketing expert system. Considering those two dimensions of future lifetime value

strongly contrasts with earlier segmentation schemes. Now, the decision maker is forced to consider the risk of a customer (group) not being as profitable as was expected or the chance a customer (group) is more profitable than was expected. As such, the segmentation scheme allows the management to have insight into the portfolio of customers in a more meaningful way and, therefore, is aided in making more optimal decisions for the company.

While we believe that this study contributes to today's literature, some shortcomings and directions for future research are given. The proposed methodology is only applied to one setting, and therefore should be validated in other cases and industries. Further research should investigate, across different datasets, the optimal choice for the τ parameter when predicting the top- x percent of valuable customers. Also with the proposed segmentation scheme, you can think of finer market segmentations. It might be fruitful to extend the proposed scheme by some variables proposed in previous research. However, we believe the proposed segmentation scheme is a valuable starting point.

Acknowledgements

The authors would like to thank the anonymous company that supplied the data to perform this research study. This research was supported by the Interuniversity Attraction Poles research network P6/03 of the Belgian Government (Belgian Science Policy).

References

- [1] Berger, P., & Nasr, N. (1998). Customer lifetime value: marketing models and applications. *Journal of Interactive Marketing*, 12(1), 49-61.
- [2] Besag, J. & Higdon, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society Series B – Statistical Methodology*, 61(4), 691-717.
- [3] Bolton, R.N., Kannan, P.K. & Bramlett, M.D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science*, 28(1), 95-108.
- [4] Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *The Journal of Human Resources*, 33(1), 88-126.
- [5] Coussement, K. & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with applications*, 34(1), 313-327.
- [6] Donkers, B., Verhoef, P. & de Jong, M. (2007). Modeling CLV: a test of competing models in the insurance industry. *Quantitative Marketing and Economics*, 5(2), 163-190.
- [7] Dotson, J.P., Retzer, J. & Allenby, G.M. (2007). Non-normal regression models for customer linkage analysis. Working paper.
- [8] Duboff, R.S. (1992). Marketing to maximize profitability. *The Journal of Business Strategy*, 13(6), 10-13.
- [9] Fader, P.S., Hardie, B.G.S. & Lee, K.L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, 42(4), 415-430.
- [10] Glady, N., Baesens, B. & Croux, C. (2008). A modified Pareto/NBD approach for predicting customer lifetime value. *Expert Systems with Applications*, forthcoming.
- [11] Gloy, B.A., Akridge, J.T. & Preckel, P.V. (1997). Customer lifetime value: An application in the rural petroleum market. *Agribusiness*, 13(3), 335-347.
- [12] Guiso, L., Halioussos, M. & Jappelli, T. (2002). *Household portfolios*. Cambridge: MIT Press.
- [13] Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N. & Sriram, N.R.S (2006). Modeling customer lifetime value. *Journal of Service Research*, 9(2), 139-155.
- [14] Hansotia, B. and Rukstales, B. (2002). Incremental value modeling. *Journal of Interactive Marketing*, 16(3), 35-46.
- [15] Hawkes, V.A. (2000). The heart of the matter: the challenge of customer lifetime value. *CRM Forum Resources*, 13, 2-10.

- [16] Hwang, H., Jung, T. & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26(2), 181-188.
- [17] Javalgi, R.G. & Dion, P. (1999). A life cycle segmentation approach to marketing financial products and services. *The Services Industries Journal*, 19(3), 74-96.
- [18] Johnson, M.D. & Selnes, F. (2004). Customer portfolio management: Toward a dynamic theory of exchange relationships. *Journal of Marketing*, 68(2), 1-17.
- [19] Kim, S.Y., Jung, T.S., Suh, E.H. & Hwang, H.S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, 31(1), 101-107.
- [20] Kim, E. & Lee, B. (2007). An economic analysis of customer selection and leverage strategies in a market where network externalities exist. *Decision Support Systems*, 44(1), 124-134.
- [21] Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.
- [22] Koenker, R. & Basset, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50.
- [23] Koenker, R. & Hallock, K.F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143-156.
- [24] Kumar, V., Lemon, K. & Parasuraman, A. (2006). Managing customers for value: An overview and research agenda. *Journal of Service Research*, 9(2), 87-94.
- [25] Malthouse, E.C. & Blattberg, R.C (2005). Can we predict customer lifetime value. *Journal of Interactive Marketing*, 19(1), 2-16.
- [26] Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7, 983-999.
- [27] Ngai, E.W.T., Li Xiu & Chau, D.C.K. (2008). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, forthcoming.
- [28] Paas, L.J., Bijmolt, T.H. & Vermunt, J.K. (2007). Acquisition patterns of financial products: A longitudinal investigation. *Journal of Economic Psychology*, 28(2), 229-249.
- [29] Pfeifer, P.E., Haskins, M.E. & Conroy, R.M. (2005). Customer lifetime value, customer profitability and the treatment of acquisition spending. *Journal of Managerial Issues*, 17(1).
- [30] Prinzie, A. & Van den Poel, D. (2007). Predicting home-appliance acquisition sequences: Markov/Markov for discrimination and survival analysis for modeling sequential information in NPTB models. *Decision Support Systems*, 44(1), 28-45.
- [31] Reinartz, W. & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67(1), 77-99.

- [32] Rossi, P.E., Allenby, G.M. & McCulloch, R.E. (2005). *Bayesian Statistics and Marketing*. New-York: Wiley.
- [33] Rossi, P.E., McCulloch, R.E. & Allenby, G.M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4), 321-340.
- [34] Rust, R. Zeithaml, V. & Lemon, K. (2000). *Driving customer equity: How customer lifetime value is reshaping corporate strategy*. New-York: The Free Press.
- [35] Schmittlein, D.C., Morrison, D.G. & Colombo, R. (1987). Counting your customers: Who are they and what will they do next. *Management Science*, 33(1), 1-24.
- [36] Venkatesan, R. & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4), 105-125.
- [37] Venkatesan, R., Kumar, V. & Bohling, T. (2007). Optimal customer relationship management using Bayesian decision theory: An application for customer selection. *Journal of Marketing Research*, 44(4), 579-594.
- [38] Verhoef, P.C. & Donkers, B. (2001). Predicting customer potential value an application in the insurance industry. *Decision Support Systems*, 32(2), 189-199.
- [39] Wübben, M. & von Wangenheim, F. (2008). Instant customer base analysis – Managerial heuristics often “get it right”. *Journal of Marketing*, forthcoming.
- [40] Yu, K. Lu, Z. & Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician*, 52(3), 331-350.
- [41] Yu, K. & Moyeed, R.A. (2001). Bayesian Quantile Regression. *Statistics and Probability Letters*, 54(4), 437-447.
- [42] Yu, K. & Zang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics: Theory & Methods*, 34(9-10), 1867-1879.
- [43] Zeithaml, V.A, Rust, R.T. & Lemon, K.N. (2001). The customer pyramid: Creating and serving profitable customers. *California Management Review*, 42(4), 118-142.

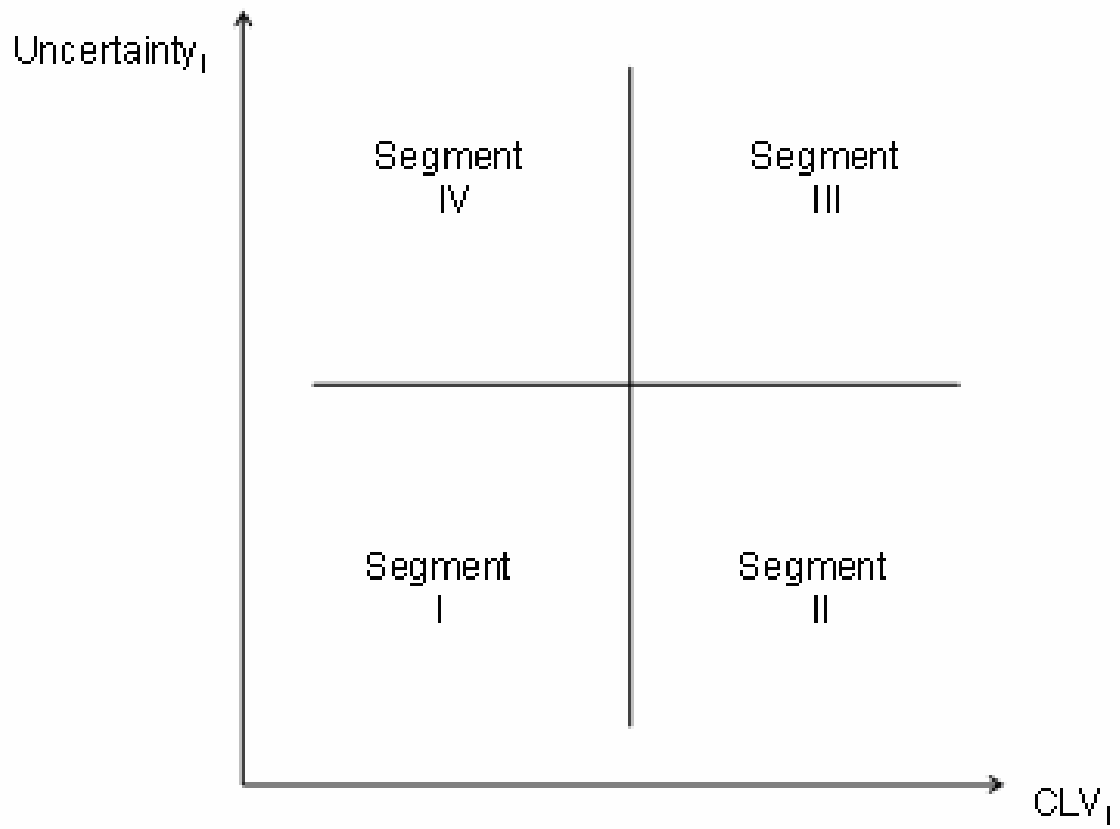


Figure 1: Segmentation with CLV and uncertainty

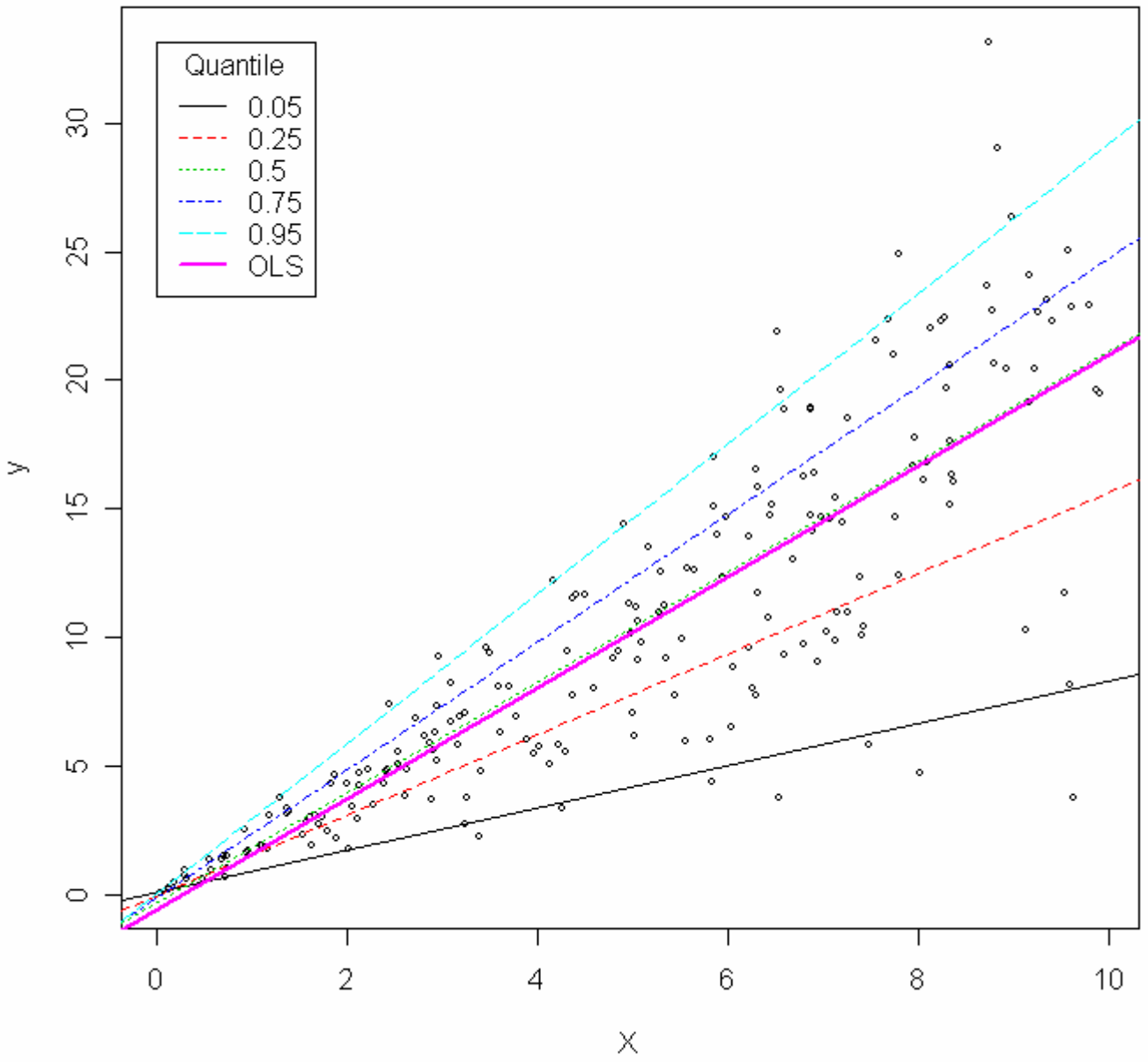


Figure 2: Illustration of Quantile Regression

Name of Variable	Description
<i>CLV</i>	The dependent variable: lifetime value based on Equations 1 and 2.
<i>Age_Max</i>	Age of the oldest household member.
<i>Nbr_FamInd</i>	Number of individuals that are part of the household.
<i>nat_nbr_Buur</i>	Number of household members with a non-Belgian nationality.
<i>Social_Class_Score</i>	A score (between 0 and 1000) indicating the social class of the neighbourhood. A high score means a higher social class.
<i>FTHB_gez</i>	A dummy variable indicating whether the household is in a given segment or not. This is: household members are under 35 and the household is a stable economic entity with a first real estate project.
<i>agent</i>	Dummy variable indicating whether the household interacts with the company through an agent or not.
<i>freq</i>	Total number of purchases ever made by all household members.
<i>nbr_insur</i>	Total number of insurance policies ever possessed by all household members.
<i>lor</i>	Number of days since the date of the first purchase of all household members.
<i>recency</i>	Number of days since the date of the last purchase of all household members.

Table 1: Overview of the variables in the analysis.

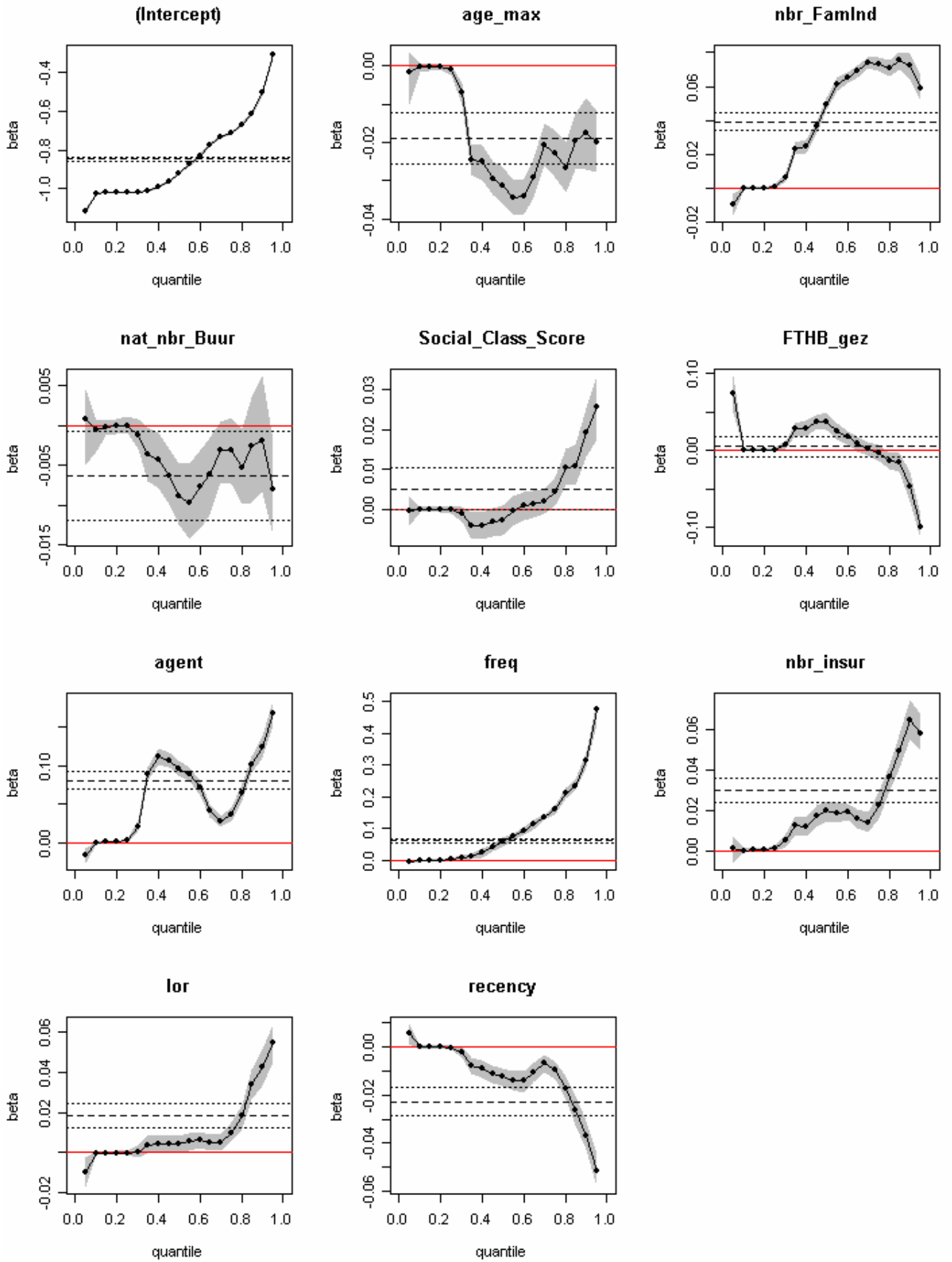


Figure 3: Quantile regression plots

	Linear Regression model	Quantile Regression model ($\tau=0.5$)	Naive mean model
hit-rate	36.39%	37.86%	26.84%

Table 2: Absolute hit-rate for the different models

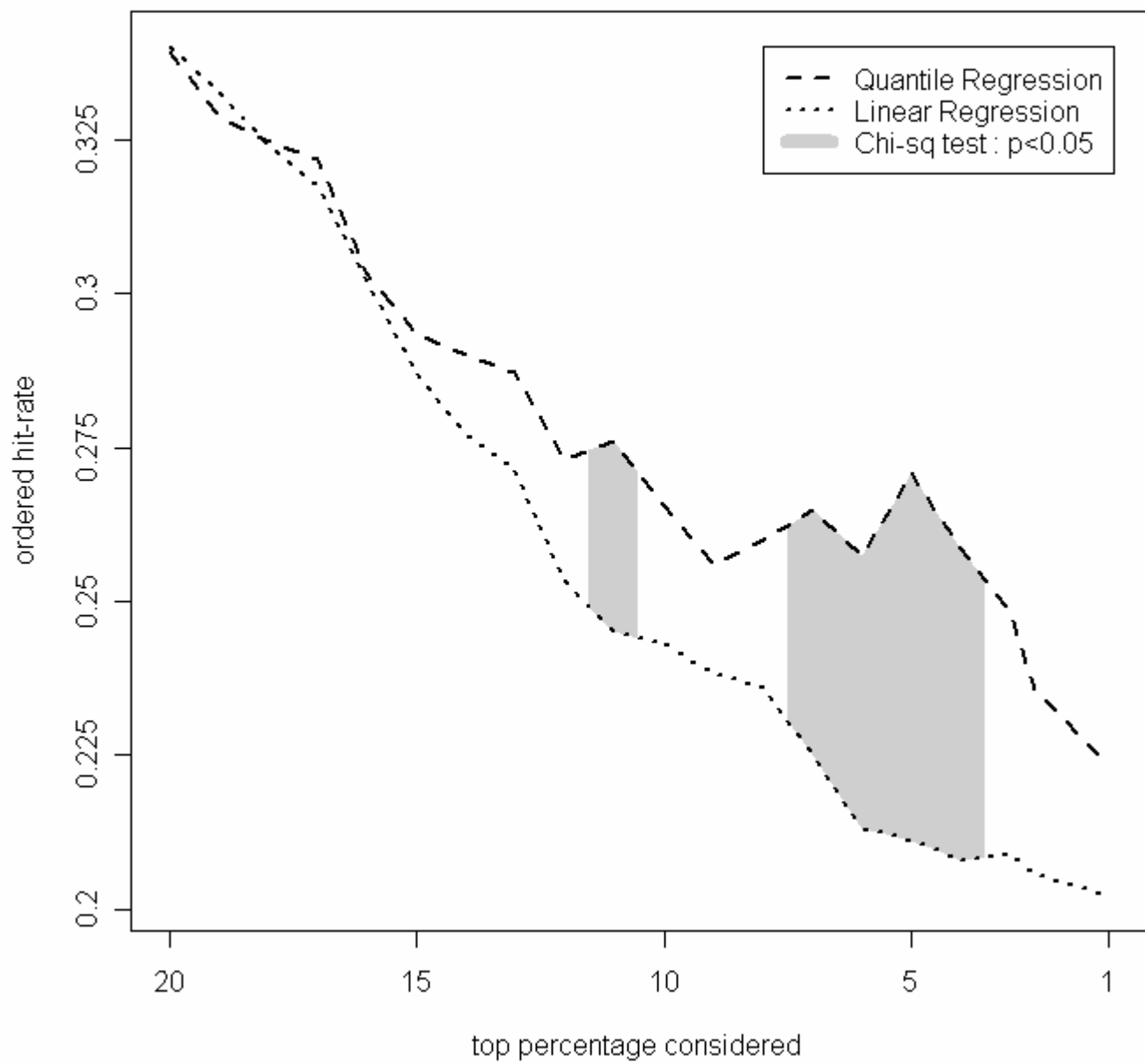


Figure 4: Ordered predictive performance

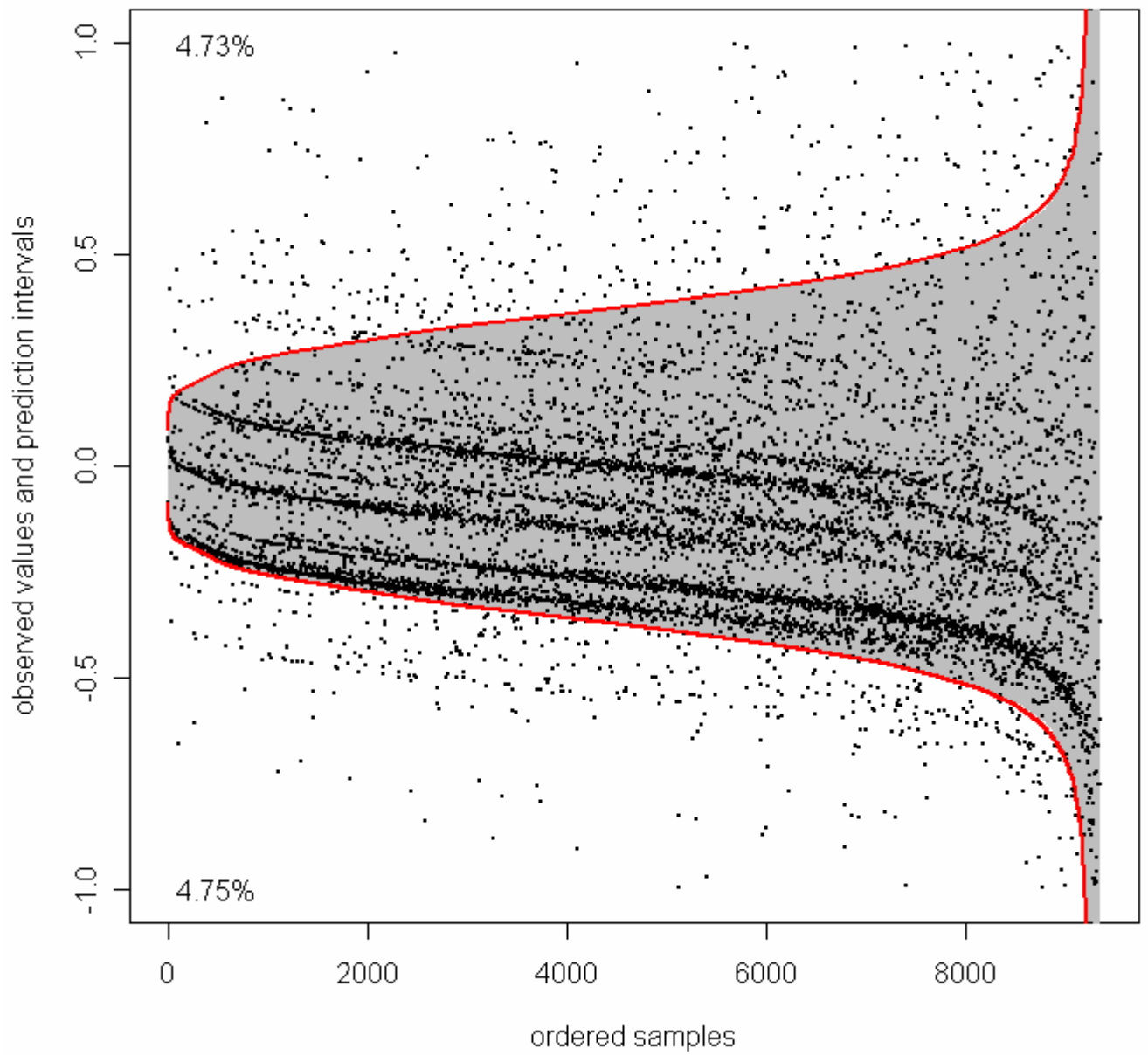


Figure 5: Graphical representation of the increasing 90% prediction intervals

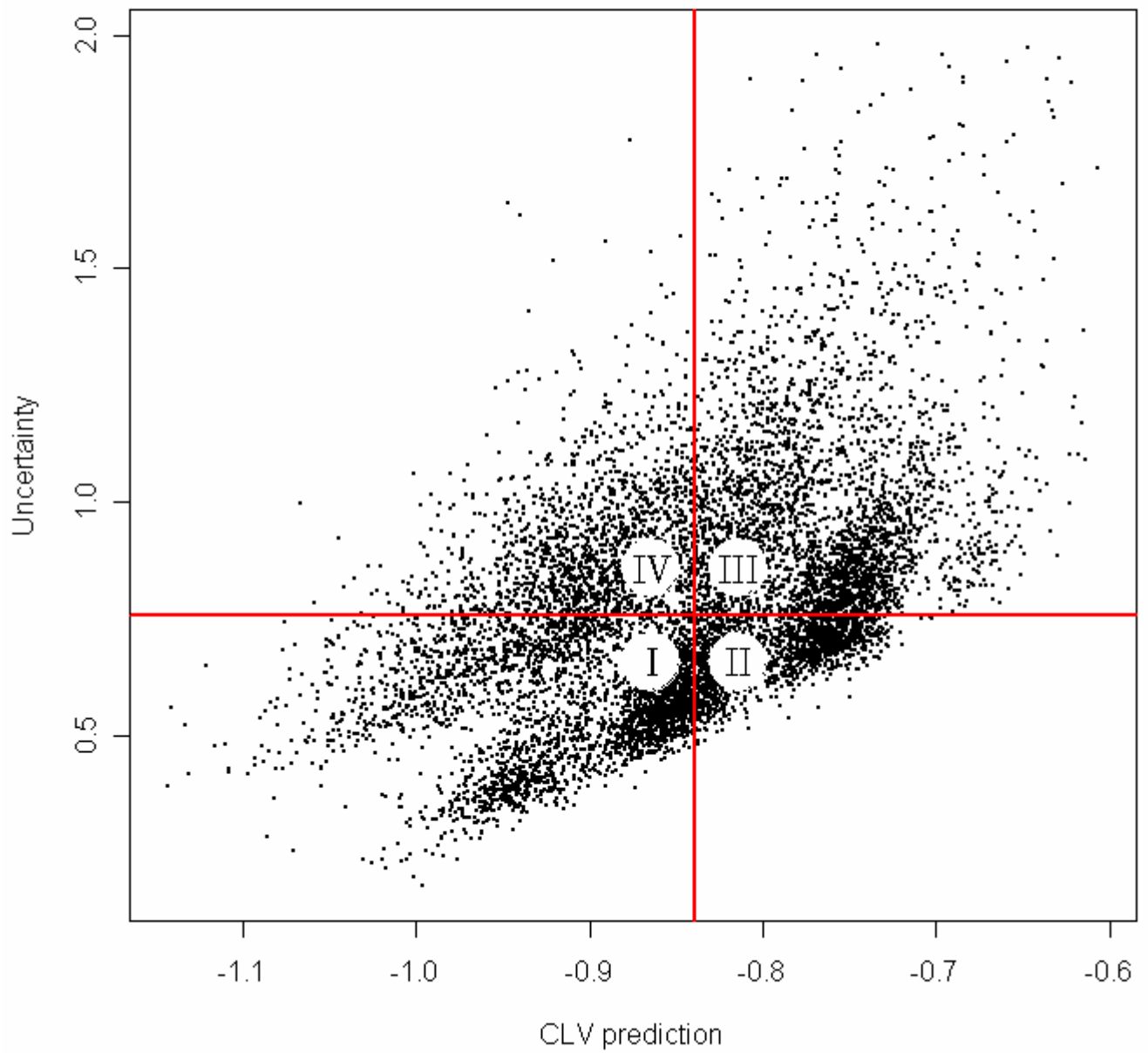


Figure 6: Result of customer segmentation using a median split.