UNIVERSITEIT GENT

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

# Enhanced Decision Support in Credit Scoring Using Bayesian Binary Quantile Regression

**Vera L. Miguéis**[1]

**Dries F. Benoit**[2]

**Dirk Van den Poel**[3]

August 2012

2012/803

[1] Vera L. Miguéis, PhD candidate, Faculty of Engineering, University of Porto, Portugal
[2] Dr. Dries F. Benoit, Faculty of Economics and Business Administration, Ghent University, Belgium.
[3] Prof. Dr. Dirk Van den Poel, Faculty of Economics and Business Administration, Ghent University, Belgium

D/2012/7012/36

# Enhanced decision support in credit scoring using Bayesian binary quantile regression

**Abstract**

Fierce competition as well as the recent financial crisis in financial and banking industries made credit scoring gain importance. An accurate estimation of credit risk helps organizations to decide whether or not to grant credit to potential customers. Many classification methods have been suggested to handle this problem in the literature. This paper proposes a model for evaluating credit risk based on binary quantile regression, using Bayesian estimation. This paper points out the distinct advantages of the latter approach: that is (i) the method provides accurate predictions of which customers may default in the future, (ii) the approach provides detailed insight into the effects of the explanatory variables on the probability of default, and (iii) the methodology is ideally suited to build a segmentation scheme of the customers in terms of risk of default and the corresponding uncertainty about the prediction. An often studied dataset from a German bank is used to show the applicability of the method proposed. The results demonstrate that the methodology can be an important tool for credit companies that want to take the credit risk of their customer fully into account.

*Key words:* Credit Scoring, Quantile regression, Classification, Bayesian estimation, Markov Chain Monte Carlo

## 1 Introduction

Credit scoring appeared around 1960 as result of the increase of computational power available to financial institutions and it drastically changed the credit process (Raymond, 2007). The existence of data warehouses made it possible to analyze large amounts of data and consequently, over the last years, financial institutions adopted quantitative analysis techniques from different fields, including statistics, operations research and computer science, to support their decisions. In the literature several definitions of credit scoring can be found. For example, Lewis (1994) defines credit scoring as a process to convert the characteristics of the applicants in numbers that are combined in order to obtain a score. This score represents the risk profile of the applicant. Verstraeten and Van den Poel (2004) refer to credit scoring as statistical methods used to classify applicants for credit in “good” versus “bad” risk classes.

Initially, credit scoring aimed at improving grant credit decisions, by making them more consistent and eliminating the subjectivity of the decision makers. More recently, credit scoring became a tool of credit process automation, such that financial intuitions can easily deal with a large number of applications. Most credit scoring models use past applicants data to estimate new applicants default risk, by assuming that credit risk is time independent. Over the years, credit scoring models included an increasing number of factors (Mays, 2004).

In the current context of crisis and regulation, financial and banking companies started to see credit risk evaluation as a critical issue. In a competitive environment, risk assessment is crucial because credit companies constantly face a potential loss of money due to defaulting customers. This makes it of great importance for companies to fully understand the process of defaulting and to be able to predict the probability of default as precisely as possible. Moreover, the number of people using credit products, such as personal loans, car loans, and credit cards from financial institutions has been generally increasing. According to Thomas et al. (2005), most of the Western adult population has some financial product from a bank or other financial institution, and some people have more than one. In this context, the ability to distinguish good customers from bad ones is crucial to ensure the sustainability of the credit business. Therefore, despite the extensive literature in this topic, efficient credit scoring models that are able to give insights about the main risk drivers and segment customers present a large potential.

This paper proposes to analyze credit scoring by means of binary quantile regression. This technique, that emerged from the work of Manski (1975), aims to fit a regression line through the conditional quantiles of the response distribution. We state that logit or probit regression provide limited information, since these methods focus on "average" effects. That is, parameter estimates of these models give information on the expected value of the latent response as a function of the predictor variables. However, in credit risk estimation, the extreme quantiles of the response distribution can be of utmost importance and should not be overlooked. Binary quantile regression provides informative analysis of the relationship between the explanatory variables and the dependent variable and this for every quantile of interest. The technique also provides an estimate of the uncertainty associated with the predicted probability of default. Using this metric, a new segmentation framework based on the probability of default versus the uncertainty associated with it can be developed.

The remainder of the paper is as follows: Section 2 includes a brief revision of the related studies present in the literature. Section 3 includes a presentation of the technique used in this paper, i.e. binary quantile regression. Section 4 introduces the measures used to evaluate the performance of the model proposed. Section 5 presents the credit scoring application based on the German credit scoring dataset. The paper finishes with a wrap-up of the findings and some managerial conclusions.

## 2 Related studies

Over the years, several approaches to deal with credit scoring have been proposed in literature. Raymond (2007); Thomas et al. (2002); Galindo and Tamayo (2000) provide summaries of the techniques used in this context. Discriminant analysis (Fisher, 1936) and logistic regression (Martin, 1977), are the most commonly used techniques. Others, such as those included in non-parametric methods class (e.g. $k$-nearest neighbor (Henley and Hand, 1996)), decision trees (Quinlan, 1992) and neural networks (Mcculloch and Pitts, 1943) have also been largely applied in the field of credit scoring. There are also some approaches that combine several techniques to create a classification model (e.g. Lee and Chen (2005); Lee et al. (2002)).

Despite the intense study of credit scoring, there is no consensus on the most appropriate classification technique to use. In fact, there are studies in which a technique is said to be more efficient than another and there are others where the same is not confirmed (e.g. Schebesch and Stecking (2003); Bellotti and Crook (2009)). Therefore, it is important to be aware of the conflicts that arise when comparing the conclusions of different studies, as stated by Baesens et al. (2003b). Nevertheless, literature (e.g. Thomas et al. (2002)) also suggests that most techniques applied in credit scoring have similar levels of performance.

From the point of view of the financial institutions the factors that can motivate the preference for a certain technique are the interpretability and the transparency (Martens et al., 2009). There are some regulators that require the institutions to indicate the reasons why they reject a credit proposal. This requirement results in the necessity for techniques to be able to identify the factors that underlie the rejections.

The above shows that at least two aspects of methods for credit scoring are very important: that is the predictive performance, as well as the insights or interpretations that are revealed by the model. In this paper, we argue that quantile regression is then an ideal choice. Quantile regression, introduced by Koenker and Bassett (1978), has shown its value in numerous studies. For example, Buchinsky (1998, 1994); Chamberlain (1994); Kahn (1998) apply this technique in the study of the dynamics of the market for labour. Arias et al. (2001) uses quantile regression to study the heterogeneity in returns to education. Engle and Manganelli (2004); Jr. and Chen (2001); Umantsev and Chernozhukov (2001) apply quantile regression on value at risk. Benoit and Van den Poel (2009) apply this technique in the analysis of customer lifetime value.

A binary version of quantile regression was introduced by Manski (1975). The relevance of this technique for the domain of credit scoring is illustrated by previous applications (i.e. Kordas (2002); Li and Miu (2010); Songfeng (2011)). However, recently it was shown that the binary quantile regression approach used in these studies has several shortcomings related to the inference and estimation of the model parameters (for further discussion, see Benoit and Van den Poel (2011)). This study applies the Bayesian method for binary quantile regression as proposed by the latter authors. By doing so, this study avoids the pitfalls related to other approached to binary quantile regression. The approach is detailed in the next section.

There are numerous studies in the literature using Bayesian approaches to model credit scoring. Antonietta and Paolo (2003) develop a Bayesian regression model to predict the credit risk of companies classified in different sectors. Maltritz and Molchanov (2008) propose a Bayesian model, in order to find the variables which are most likely to determine country default risk in emerging markets. The collection provided by Bocker (2010) also includes several studies on Bayesian credit risk modeling. For example, Jacobs and Kiefer (2010) consists of a step-by-step guide to Bayesian analysis in the default setting, including details on elicitation of expert information.

## 3 Binary quantile regression

The standard linear regression model describes how the mean of the dependent variable $y$ varies according to the explanatory variables vector $x$. However, in many cases, the effects of the explanatory variables on the dependent variable are not constant, but rather change across the values of the dependent variable, according to a distribution function. This heterogeneity is not captured when limiting the analysis to simple mean regression.

In this context, Koenker and Bassett (1978) extend the classical regression model, that focuses on the mean estimation, in a regression model that estimates the conditional effects of the entire response distribution. The $\tau$-th quantile of a distribution, $\tau$ in (0,1), represents the value such that there is $100\tau\%$ of the observations concentrated on the left side of the distribution. Such that, quantile 0.5 represents the median of the distribution.

Manski (1975) introduced median and quantile regression in a classification context. Consider for a certain instance $i$, the dependent variable of a binary regression as:

$$\begin{aligned} y_i^* &= x_i^{'}\beta_\tau + u_i \\ y_i &= 1 \text{ if } y_i^* \geq 0, y_i = 0 \text{ otherwise} \end{aligned} \tag{1}$$

$y_i^*$ is a latent continuous variable that allows to determine the dependent binary variable $y_i$, $x_i$ is a $1 \times k$ vector of explanatory variables, $\beta_\tau$ is a $k \times 1$ vector of unknown parameters to be estimated for different values of $\tau$ and $u_i$ is a random error term that is independently and identically distributed. In the original formulation, the only requirement of the distribution of $u$ is that the $\tau$-th quantile equals 0. As a result, the $\tau$-th quantile of $y_i^*$ on the explanatory vector $x$ can be defined as:

$$Q_\tau(y_i^*|x_i) = x_i^{'}\beta_\tau \tag{2}$$

Kordas (2006) proposed a method for making probabilistic predictions with the binary quantile regression model. By varying the value of $\tau$ from (0 to 1) it is possible to obtain the predicted distribution of $y^*$ given $x$ and consequently it is possible to estimate the probability of $y$ take value 0 or 1, corresponding to the class of non-defaulters and defaulters, respectively. For each instance, the probability that $y$ takes value 0 is the smallest quantile level for which the corresponding quantile is greater or equal to zero and the probability of $y$ take value 1 is the complement to one (Kordas, 2006).

Naturally, for certain instances, the estimation of the dependent variable is more accurate than for other instances. Quantile regression, in opposition to the standard mean regression techniques, can provide principled information about the dispersion around the estimations (Meinshausen, 2006). For example, a 90% prediction interval for the value of $y_i^*$ is given by:

$$I = [Q_{0.05}(y_i^*|x_i), Q_{0.95}(y_i^*|x_i)] \tag{3}$$

The width of this prediction interval can be considered a measure of the uncertainty associated with the estimation of $y_i^*$.

In order to estimate the binary quantile regression parameters $\beta_\tau$, we use the approach presented by Benoit and Van den Poel (2011). They adopt a Bayesian approach to infer the conditional posterior distribution of $\beta_\tau$. A conjugate prior distribution for the model parameters is not available (Yu and Zhang, 2005). The normal distribution is used as prior, however other choices could be made too. The authors continue by placing an asymmetric Laplace density on the latent variable $y^*$. The joint posterior density of the unobservables $\beta$ and $y_*$ given the data $y$ and the quantile of interest $\tau$ is then given by:

$$\pi(\beta, y^*|y, \tau) \propto \pi(\beta) \prod_{i=1}^{n} \{I(y^* > 0)I(y_i = 1) + I(y^* \leqslant 0)I(y_i = 0)\} \, ALD(y_i^*; \beta, 1, \tau) \qquad (4)$$

This posterior is not of known form and thus direct sampling is not possible. However Markov Chain Monte Carlo (MCMC) methods (Chib and Greenberg, 1995) can be used to sample from this distribution. Conditional on $\beta$, the posterior of $y^*$ becomes the truncated asymmetric Laplace distribution. The conditional distribution of $\beta$ is unknown and as a result Metropolis-Hastings step is required. The resulting MCMC is of the form Metropolis-Hastings within Gibbs. See Benoit and Van den Poel (2011) for more details on the sampler.

It is important to note that the parameter estimates of Bayesian methods are interpreted differently compared to frequentist methods. A comparison of both viewpoints is beyond the scope of this article. However, a detailed discussion on the two approaches can be found, for example, in Bayarri and Berger (2004).

## 4 Performance measures

In order to measure the performance of the predictions we compute the well-known receiver operating characteristic curve (ROC) and we analyze the area under that curve (AUC) (Bradley, 1997). ROC curve plots the true positives (i.e. in this case, the defaulters that were classified as defaulters) versus the false positives (i.e. the non-defaulters that were classified as defaulters), as a discrimination threshold is varied. This threshold is the minimum probability at which the dependent variable takes value 1 (i.e. default). An AUC close to 1.0 means that the model has perfect discrimination, while an AUC close to 0.5 suggests poor discrimination. Moreover, we also use the percentage of cases correctly classified (PCC). (Quinlan, 1986), also known as accuracy, as an evaluation metric of the classification model. All cases having an estimated probability above a certain discrimination threshold defined by the analyst, are considered to have $y = 1$ and all cases having a lower probability are considered to have $y = 0$. Then, from the comparison between the predictions and the observed values, the ratio between the number of correctly classified cases and the total number of cases is computed. The PCC obtained should exceed the PCC obtained when using the naive model. The naive model is defined as a model which always predicts the most common class. Both PCC and AUC have proven their value in related domains for binary classification (Verstraeten and Van den Poel, 2004).

To assess model performance, the reference dataset is split into multiple training and testing subsets. Literature recommends k-fold stratified cross validation as an appropriate method for evaluating classification techniques (Ho, 2002). According to Witten and Eibe (2005), ten is the most appropriate number of subsets or folds to get good estimation performance. In this case, the initial data is randomly partitioned into 10 mutually exclusive folds of approximately equal size and the model is trained on nine datasets and tested on one. This process is repeated ten times and usually the performance measures are the average of the measures obtained in each iteration. The folds are stratified, so that the class distribution of the samples in each fold is approximately the same as that in the initial data.

Note that a full Bayesian approach would undertake model assessment by comparing the posterior probabilities of competing Bayesian models. However, in this paper we have adopted a "pragmatic" Bayesian point of view (Alqallaf and Gustafson, 2001). This is, cross-validation is used to assess the model performance. This is not a pure Bayesian approach, but, as Alqallaf and Gustafson (2001) point out, by doing so it makes comparison with frequentist methods easier. Moreover, cross-validation also implies that model performance is checked on unseen data.

## 5 Credit scoring application

In this section, we present the real data used for the empirical analysis, the predictive performance of the model and the insights binary quantile regression provides about the relationship between the explanatory variables. This section also introduces a segmentation framework for the credit applicants. The estimation of the model parameters was done using the bayesQR R-package (Benoit et al., 2011). Since no external or historical information about the parameters was present, vague prior distributions were placed on the model parameters, i.e. $\pi(\beta) \sim Normal(0, 100)$.

### *5.1 Data*

The data used in this study is the German credit dataset, publicly available at Asuncion and Newman (2007) and was used in other studies such as Huang et al. (2007) and West et al. (2005). This dataset consists of 700 examples of creditworthy applicants and 300 examples of applicants who defaulted. For each instance, the dataset includes 24 input variables that describe 19 attributes that characterize the applicants (with 4 nominal attributes transformed to dummy variables). These explanatory variables, shown in Table A1 of the Appendix, include demographic characteristics of customers (e.g., *Personal status and gender*, *Age in years*), credit details (e.g., *Duration of credit (months)* and *Credit amount*), customers' financial standing (e.g., *Average balance in savings account*, *Credit history*) and employment (e.g., *Nature of job*, *Present employment since*). The dependent variable reveals default (value=1) or non-default (value=0).

### *5.2 Prediction and evaluation*

The performance of credit scoring models is obviously of utmost importance in financial and banking industries. To evaluate the performance of the binary quantile regression model proposed, we first estimate the regression for nineteen different quantile levels ($\tau = 0.05, ..., 0.95$ by 0.05). Consequently we compute the probability of default, according to the procedure explained in Section 3. Using the values obtained, we compute the AUC. The average accuracy obtained through a 10 cross validation is 0.77, with a standard deviation of 0.06. This means that the proposed model has good discrimination power since it significantly exceeds the null-model benchmark of 0.5. The average percentage of correctly classified instances is 76.2%, with a standard deviation of 5.1%, when using a threshold of 0.5. This is higher than the accuracy rate obtained when using a naive model, i.e. 70%, resulting from the classification of all customers as non-defaulters.

As stated before, the German credit data is used by several authors to build models in a credit scoring context. However, the results obtained in those studies should be compared with care with the results reported in this paper because of model assumptions and validation method. For example, Baesens et al. (2003a) used various methods, e.g. neural networks and decision trees, for building the credit scoring model. The most accurate model turned out to be the pruned neural network and had a PCC of 77.8%. However, this study does not use a cross-validation method to evaluate the performance of the models. Xiao et al. (2006) also applies several methods to classify credit applicants using the German credit data. This study does use cross-validation. The most accurate model, i.e. support vector machines with sigmoid kernel, presents a PCC of 77.2%. Although the evaluation method is the same as the one used in this paper, we don't have information regarding the threshold that was used. This choice can significantly influence the resulting PCC. However, given the results of the current study and the results reported in the previous studies, we conclude that the model proposed can compete with the state-of-the-art models suggested in the literature.

### *5.3 Explanatory variables effects*

Often used parametric models, such as logit or probit, give insight into the effect size of the mean of the response distribution. However, with binary quantile regression, it is possible to get a more thorough view on the effect of the explanatory variables. To do this, we analyzed the regression parameters for the quantile levels $\tau = 0.05, ..., 0.95$ by 0.05.

It is important to note that before estimating the regression model, we considered the possible multicollinearity issues. There are a number of approaches to deal with multi-collinearity. One popular check is the variance inflation factor (VIF) (see Bajpai (2009)). A value of VIF greater than 10 is an indication that multicollinearity may be causing problems in estimations (Neter et al., 1996; Myers, 2000). VIF statistics for the independent variables considered in this study indicate VIF ranging from 1.1 to 3.3. This is within the acceptable range and thus multicollinearity is not an important issue for the current dataset.

To analyze the regression parameters we computed a 90% pointwise Bayesian credible interval from the marginal posterior distributions of each parameter. The results showed that, for 16 variables, the confidence intervals of the regression parameters overlap the value of zero on practically all quantile levels. Therefore, we conclude that these variables are not important for the analysis. Most of these variables are demographic variables and variables concerned with the employment. Moreover, variables related to credit details, such as *Credit amount* and *Application has other debtors or guarantors:Co-applicant*, and customers' financial standing, such as *Number of existing credits at this bank* and *Other installment plans*, also seem not to influence credit score estimation.

Figure 1 depicts a summary of the quantile regression parameters obtained for the 8 variables that are relevant on most quantile levels. The solid line with filled dots represents the point estimates of the regression coefficients for the different quantile levels ($\tau = 0.05, ..., 0.95$ by 0.05). The shaded area represents the 90% pointwise credible intervals obtained from the marginal posterior distribution of the different regression parameters.

By analyzing Figure 1 we can conclude there are some variables whose impact is negative at each quantile level, while there are others whose impact is positive. None of the relevant variables

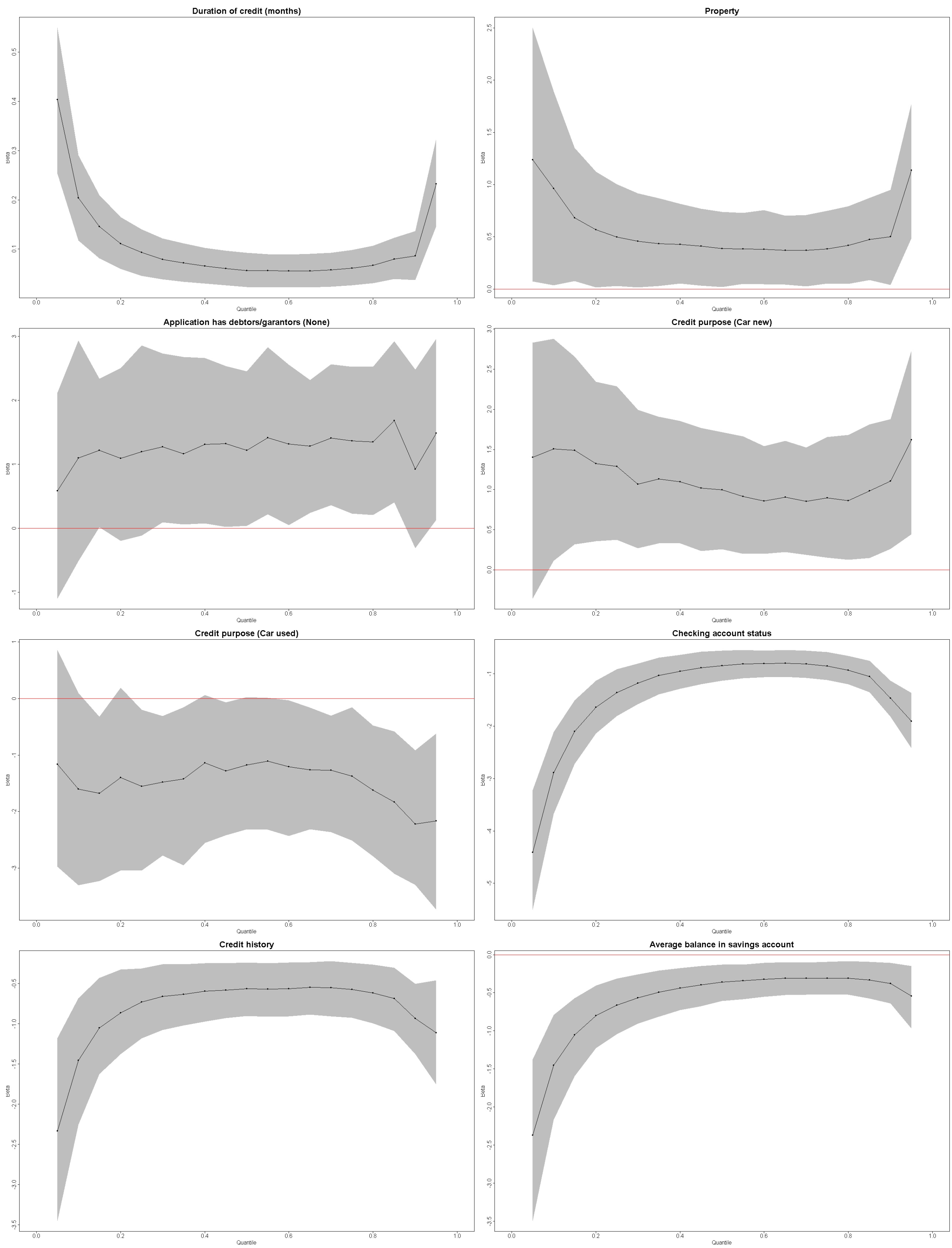


Fig. 1. Regression parameters.

have opposite effects for lower versus higher quantiles. However, the impact of each variable on credit risk seems to be different over quantile levels. This reinforces the supremacy of quantile regression over other techniques used in this context, which assume a constant effect of the explanatory variables at different points of the independent variable distribution. The variable *Duration of credit (months)* has a positive impact on the risk of failure. This impact is higher in the extreme high and low quantiles, than in the middle quantiles. This means that analyzing only the mean effect would give the researcher an overestimation of the typical effect of this variable. The variable *Property* also has a positive relationship with credit risk, suggesting that customers with less valuable properties (see Table A1) are those with higher promptness to default. This tendency is, once again, more evident for both low and high extreme quantiles. For most quantiles, the variable *Application has other debtors or guarantors:Co-applicant:None* positively influences the credit risk. This suggests that people who apply alone for a credit have higher probability of not succeeding in repaying it than people who have support from other people. This idea is not valid for low quantile levels and one high quantile level since the relationship seems to be relevant. The variable *Credit purpose: New car* also presents a positive impact on the dependent variable analyzed. However, this impact is not relevant for creditworthy applicants (i.e., low quantiles). The effect of this variable is practically stable over quantiles. This means that for this variable, the additional insights are limited compared to the insights from logit or probit models. In contrast, the variable *Credit purpose: Used car* has a negative impact. This is more pronounced for high quantiles. As expected, both *Customer account status* and *Average balance in savings account* present a negative relationship with credit risk, revealing that people with more money in the accounts are creditworthy applicants. This tendency is more outspoken for low quantile levels, i.e. customers with better credit quality given the set of covariates. Concerning the variable *Credit history*, it is interesting to observe that it also has a negative impact, suggesting that people with a compromising history are less prone to default. It may reveal that the possible inconvenients arising from the past credit processes made them averse to failure.

### *5.4 Customer segmentation*

As discussed in Section 3, quantile regression allows to compute uncertainty estimates. Equation (3) allows to calculate uncertainty intervals, whose amplitude can be considered a measure of the uncertainty associated with the estimation of the independent variable. We propose to use the probability of default as well as this uncertainty measure to segment customers. Each group of customers represents a different level of credit risk and uncertainty.

For illustration purposes consider Figure 2 that depicts the estimates of the probability of default for each customer and the corresponding uncertainty. We segment customers in 4 groups, by considering a split between customers whose probability of default is lower or higher than 50%, and a split between customers whose estimation uncertainty is lower or higher than the empirical median. However, it is important to note that according to the risk aversion of the financial companies or other managerial considerations, other choices for cut-off values for the segments can be more appropriate.

From the analysis of Figure 2, we conclude that for each probability level there are several levels of uncertainty. For example, for a probability of default of 5% the uncertainty estimation ranges from 16.8 to 50.3. Segment I includes customers whose probability of default is relatively high (i.e. higher than 50%), while the uncertainty is relatively low (i.e. less than the median), what corresponds to 19.4% of the customers considered in the study. This is a group of customers who can be considered undesired for the company. Segment II includes only 1.2% of the customers

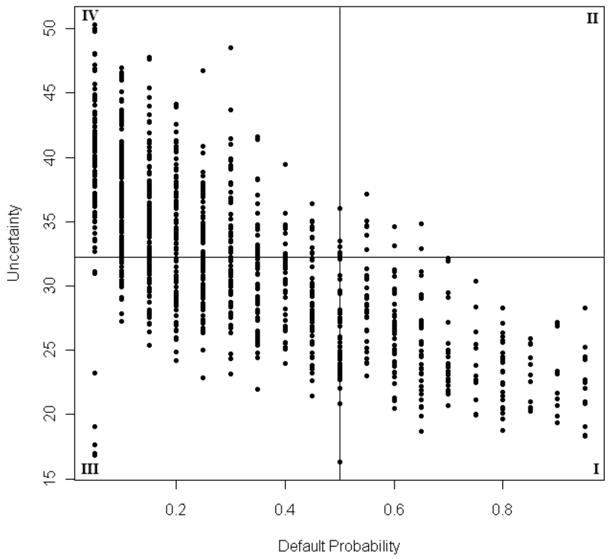


Fig. 2. Customer segmentation.

analyzed. Despite having a relative high probability of default (i.e. higher than 50%), there is more uncertainty associated with this estimative (i.e. higher or equal to the median). Such that, company should be aware that this is a more risky group of customers than Segment I. More information may be collected in order to get more insights about their promptness to default and thus decrease the uncertainty around the prediction. Or, alternatively, a higher interest rate might be proposed to these customers to account for the greater risk for the credit supplier. Segment III includes customers whose probability of default is less or equal to 0.5 and whose uncertainty is less than the median, which corresponds to 30.5% of the customers analyzed. These customers are the most interesting customers for the company, since there is relative certainty that the credit will be paid back to the credit supplier. Finally, Segment IV is the biggest segment with 48.8% of the customers being part of it. These customers present a relatively low probability of default (i.e. less or equal to 0.5) but this is rather uncertain (i.e. higher or equal to the median). This representative segment is a critical group due to the uncertainty associated with the estimates. Managers can use extra information to get deeper insights into the risk of default of this group of customers. Companies can also initiate a close follow-up in the case of granting credit to customers in this segment.

## 6 Conclusion

This paper proposes to assess credit risk based on binary quantile regression in order to support financial companies in their credit granting process. The applicability of the method is shown on the often studied German credit dataset.

The paper adds to the existing literature in three distinct ways. First, by introducing a technique that accurately predicts customers default. An important finding is that the current method can compete with other sophisticated methods used in this context. Second, the methodology proposed is not a black-box method such as random forests or support vector machines. Moreover, the method gives additional insights into the effects of the explanatory variables that are totally missed with other existing methods such as the popular logit or probit models. Finally, the method provides the necessary input for segmenting the customers in terms of risk of default and corresponding uncertainty. This type of segmentation is really valuable, because it acknowledges the notion that potential customers with equal probability of default can have different uncertainty about this default probability and thus are not equally valuable to the credit company. The segmentation framework can also guide other decisions. For instance, it can help companies to decide whether or not to offer other financial products to the customer, or whether to give them unsolicited increases in their credit limit. The customer segmentation can also help financial and banking companies to decide whether to offer customers better features on their credit product so as to retain them and prevent them from moving to competitors.

## References


Alqallaf, F. and Gustafson, P. (2001). On cross-validation of bayesian models. *Canadian Journal of Statistics-Revue Canadienne De Statistique*, 29(2):333–340. WOS:000169424700011.

Antonietta, M. and Paolo, T. (2003). Bayesian estimate of credit risk via MCMC with delayed rejection. Economics and quantitative methods, Department of Economics, University of Insubria.

Arias, O., Sosa-Escudero, W., and Hallock, K. F. (2001). Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics*, 26(1):7–40.

Asuncion, A. and Newman, D. (2007). UCI machine learning repository. http://archive.ics.uci.edu/ml/.

Baesens, B., Setiono, R., Mues, C., and Vanthienen, J. (2003a). Using neural network rule extraction and decision tables for Credit-Risk evaluation. *Management Science*, 49:312–329.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003b). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635.

Bajpai, N. (2009). *Business Statistics*. Pearson, India.

Bayarri, M. and Berger, J. (2004). The interplay of bayesian and frequentist analysis. *Statistical Science*, 19(1):58–80. WOS:000223012900009.

Bellotti, T. and Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36:3302–3308.

Benoit, D. F., Al-Hamzawi, R., Yu, K., and Van den Poel, D. (2011). *bayesQR: Bayesian quantile regression*. R package version 1.3.

Benoit, D. F. and Van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. *Expert Systems with Applications*, 36(7):10475–10484.

Benoit, D. F. and Van den Poel, D. (2011). Binary quantile regression: a bayesian approach based on the asymmetric laplace distribution. *Journal of Applied Econometrics*, Forthcoming (doi:10.1002/jae.1216).

Bocker, K. (2010). *Rethinking Risk Measurement and Reporting: Volume II.* Risk Books, London.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145—1159.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2):405–58.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, 13(1):1–30.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In Sims, C., editor, *Advances in econometrics - Sixth World Congress*, volume 1. Cambridge University Press, Cambridge.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22:367–381.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals Eugen.*, 7:179–188.

Galindo, J. and Tamayo, P. (2000). Credit risk assessment using statistical and MachineLearning: basic methodology and risk modeling applications. *Comput. Econ.*, 15(1-2):107143.

Henley, W. E. and Hand, D. J. (1996). A k-Nearest-Neighbour classifier for assessing consumer credit risk. *The Statistician*, 45(1):77.

Ho, T. (2002). Rule induction in constructing Knowledge-Based decision support. In Kersten, G., Mikolajuk, Z., and Yeh, A., editors, *Decision Support Systems for Sustainable Development*, pages 263–276. Kluwer Academic Publishers, USA.

Huang, C., Chen, M., and Wang, C. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 33(4):847–856.

Jacobs, M. and Kiefer, N. (2010). The bayesian approach to default risk: A guide. In K., B., editor, *Rethinking Risk Measurement and Reporting: Volume II*, pages 319–334. Risk Books, London.

Jr., G. W. B. and Chen, H. (2001). Portfolio style: Return-based attribution using quantile regression. *Empirical Economics*, 26(1):293–305.

Kahn, L. M. (1998). Collective bargaining and the interindustry wage structure: International evidence. *Economica*, 65(260):507–34.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Kordas, G. (2002). Credit scoring using binary quantile regression. In Dodge, Y., editor, *Statistical Data Analysis Based on the L1-Norm and Related Methods*. Birkhuser Basel.

Kordas, G. (2006). Smoothed binary regression quantiles. *Journal of Applied Econometrics*, 21(3):387–407.

Lee, T. and Chen, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4):743–752.

Lee, T., Chiu, C., Lu, C., and Chen, I. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, 23(3):245–254.

Lewis, E. M. (1994). *An introduction to credit scoring*. Athena Press, 2nd edition.

Li, M. L. and Miu, P. (2010). A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance*, 17(4):818–833.

Maltritz, D. and Molchanov, A. (2008). Economic determinants of country credit risk: A bayesian approach. In *Proceedings of the 12th New Zealand Finance Colloquium*, Massey University, Palmerston North, NZ. Massey University.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3):205–228.

Martens, D., Van Gestel, T., De Backer, M., Haesen, R., Vanthienen, J., and Baesens, B. (2009). Credit rating prediction using ant colony optimization. *Journal of the Operational Research Society*, 61(4):561–573.

Martin, D. (1977). Early warning of bank failure : A logit regression approach. *Journal of Banking & Finance*, 1(3):249–276.

Mays, F. E. (2004). *Credit scoring for risk managers: the handbook for lenders*. South Western Educational Publishing, United States of America, 1st edition.

Mcculloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysic*, 5:115–133.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, 7:983—999.

Myers, R. H. (2000). *Classical and Modern Regression with Applications*. Duxbury Press, Boston, 2 edition.

Neter, J., Kutner, M., Wasserman, W., and Nachtsheim, C. (1996). *Applied Linear Statistical Models*. McGraw-Hill/Irwin, New York, 4 edition.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.

Quinlan, J. R. (1992). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., California.

Raymond, A. (2007). *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press, United States of America, 1st edition.

Schebesch, K. and Stecking, R. (2003). Support vector machines for credit scoring: Extension to non standard cases. In Baier, D. and Wernecke, K., editors, *Innovations in Classification, Data Science, and Information Systems*, Cottbus.

Songfeng, Z. (2011). QBoost: predicting quantiles with boosting for regression and binary classification. *Expert Systems with Applications*, Forthcoming (doi:10.1016/j.eswa.2011.06.060)(0).

Thomas, L. C., Edelman, D. B., and Crook, J. N. (2002). *Credit Scoring & Its Applications*. Society for Industrial Mathematics, Philadelphia, 1st edition.

Thomas, L. C., Oliver, R. W., and Hand, D. J. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9):1006–1015.

Umantsev, L. and Chernozhukov, V. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292.

Verstraeten, G. and Van den Poel, D. (2004). The impact of sample bias on consumer credit scoring performance and profitability. *Journal of the Operational Research Society*, 56(8):981–992.

West, D., Dellana, S., and Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10):2543–2559.

Witten, I. H. and Eibe, F. (2005). *Data Mining*. Morgan Kaufmann, San Francisco.

Xiao, W., Zhao, Q., and Fei, Q. (2006). A comparative study of data mining methods in consumer loans credit scoring management. *Journal of Systems Science and Systems Engineering*, 15:419–435.

Yu, K. and Zhang, J. (2005). A Three-Parameter asymmetric laplace distribution and its extension. *Communications in Statistics: Theory and Methods*, 34:1867–1879.

# APPENDIX

Table A1: Dataset variables.

| Attribute | Variables | Type | Name | Description |
|---|---|---|---|---|
| 1 | 1 | Qualitative | Checking account status | 0 : < 0 DM<br>1: 0 < ...< 200 DM<br>2 : >= 200 DM<br>3: no checking account |
| 2 | 2 | Quantitative | Duration of credit (months) | |
| 3 | 3 | Qualitative | Credit history | 0: no credits taken<br>1: all credits at this bank paid back duly<br>2: existing credits paid back duly till now<br>3: delay in paying off in the past<br>4: critical account |
| 4 | 4 | Quantitative | Credit amount | |
| 5 | 5 | Qualitative | Average balance in savings account | 0 : < 100 DM<br>1 : 100<= ... < 500 DM<br>2 : 500<= ... < 1000 DM<br>3 : >=1000 DM<br>4 : unknown/ no savings account |
| 6 | 6 | Qualitative | Present employment since | 0 : unemployed<br>1: < 1 year<br>2 : 1 <= ... < 4 years<br>3 : 4 <=... < 7 years<br>4 : >= 7 years |
| 7 | 7 | Qualitative | Personal status and gender | 0: male and divorced/separated<br>1: female and divorced/separated/married<br>3: male and single<br>4: male and married/widowed<br>5: female and single |
| 8 | 8 | Qualitative | Present residence since | 0: <= 1 year<br>1: 1< ... <=2 years<br>2: 2< ... <=3 years<br>3:>4years |
| 9 | 9 | Qualitative | Property | 0: real state<br>1: building society savings agreement/life insurance (not 0)<br>2: car or other<br>3: unknown/no property |
| 10 | 10 | Quantitative | Age (years) | |
| 11 | 11 | Quantitative | Number of existing credits at this bank | |
| 12 | 12 | Qualitative | Other installment plans | 0: bank<br>1: stores<br>3: none |
| 13 | 13 | Quantitative | Number of people for whom liable to provide maintenance | |
| 14 | 14 | Qualitative | Applicant has phone in his or her name | 1: yes |

Continued on next page

**Table A1 – continued from previous page**

| Attribute | Variables | Type | Name | Description |
|---|---|---|---|---|
| | | | | 2: no |
| 15 | 15 | Qualitative | Foreign worker | 1: yes<br>2: no |
| 16 | 16-17 | Dummy | Credit purpose | Car new, car used |
| 17 | 18-19 | Dummy | Application has other debtors or guarantors | None, co-applicant |
| 18 | 20-21 | Dummy | Nature of house | Rented house, own house |
| 19 | 22-24 | Dummy | Nature of job | Unemplyed or unskilled-non resident, unskilled-resident, sikked or official |